PZ006 PCT

# PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)


(51) International Patent Classification 6 : C07K 1/00, C07H 21/04 — A1

(11) International Publication Number: WO 98/54206

(43) International Publication Date: 3 December 1998 (03.12.98)

(21) International Application Number: PCT/US98/10868

(22) International Filing Date: 28 May 1998 (28.05.98)

(30) Priority Data:

| | | |
|---|---|---|
| 60/044,039 | 30 May 1997 (30.05.97) | US |
| 60/048,093 | 30 May 1997 (30.05.97) | US |
| 60/048,190 | 30 May 1997 (30.05.97) | US |
| 60/050,935 | 30 May 1997 (30.05.97) | US |
| 60/048,101 | 30 May 1997 (30.05.97) | US |
| 60/048,356 | 30 May 1997 (30.05.97) | US |
| 60/056,250 | 29 August 1997 (29.08.97) | US |
| 60/056,296 | 29 August 1997 (29.08.97) | US |
| 60/056,293 | 29 August 1997 (29.08.97) | US |



(71) Applicant *(for all designated States except US)*: HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: RUBEN, Steven, M. [US/US]; 18528 Heritage Hills Drive, Olney, MD 20832 (US). ROSEN, Craig, A. [US/US]; 22400 Rolling Hill Road, Laytonsville, MD 20882 (US). CARTER, Kenneth, C. [US/US]; 11601 Brandy Hall Lane, North Potomac, MD 20878 (US). DILLON, Patrick, J. [US/US]; 1055 Snipe Court, Carlsbad, CA 92009 (US). ENDRESS, Gregory, A. [US/US]; 9729 Clagett Farm Drive, Potomac, MD 20854 (US). YU, Guo-Liang [CN/US]; 13524 Straw Bale Lane, Darnestown, MD 20878 (US). NI, Jian [CN/US]; 5502 Manorfield Road, Rockville, MD 20853 (US). FENG, Ping [CN/US]; 4 Relda Court, Gaithersburg, MD 20878 (US).

(74) Agents: BROOKES, A., Anders et al.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 20850 (US).



(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

**Published**
*With international search report.*
*Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.*


(54) Title: 32 HUMAN SECRETED PROTEINS

(57) Abstract


The present invention relates to 32 novel human secreted proteins and isolated nucleic acids containing the coding regions of the genes encoding such proteins. Also provided are vectors, host cells, antibodies, and recombinant methods for producing human secreted proteins. The invention further relates to diagnostic and therapeutic methods useful for diagnosing and treating disorders related to these novel human secreted proteins.

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | |
|---|---|
| AL | Albania |
| AM | Armenia |
| AT | Austria |
| AU | Australia |
| AZ | Azerbaijan |
| BA | Bosnia and Herzegovina |
| BB | Barbados |
| BE | Belgium |
| BF | Burkina Faso |
| BG | Bulgaria |
| BJ | Benin |
| BR | Brazil |
| BY | Belarus |
| CA | Canada |
| CF | Central African Republic |
| CG | Congo |
| CH | Switzerland |
| CI | Côte d'Ivoire |
| CM | Cameroon |
| CN | China |
| CU | Cuba |
| CZ | Czech Republic |
| DE | Germany |
| DK | Denmark |
| EE | Estonia |
| ES | Spain |
| FI | Finland |
| FR | France |
| GA | Gabon |
| GB | United Kingdom |
| GE | Georgia |
| GH | Ghana |
| GN | Guinea |
| GR | Greece |
| HU | Hungary |
| IE | Ireland |
| IL | Israel |
| IS | Iceland |
| IT | Italy |
| JP | Japan |
| KE | Kenya |
| KG | Kyrgyzstan |
| KP | Democratic People's Republic of Korea |
| KR | Republic of Korea |
| KZ | Kazakstan |
| LC | Saint Lucia |
| LI | Liechtenstein |
| LK | Sri Lanka |
| LR | Liberia |
| LS | Lesotho |
| LT | Lithuania |
| LU | Luxembourg |
| LV | Latvia |
| MC | Monaco |
| MD | Republic of Moldova |
| MG | Madagascar |
| MK | The former Yugoslav Republic of Macedonia |
| ML | Mali |
| MN | Mongolia |
| MR | Mauritania |
| MW | Malawi |
| MX | Mexico |
| NE | Niger |
| NL | Netherlands |
| NO | Norway |
| NZ | New Zealand |
| PL | Poland |
| PT | Portugal |
| RO | Romania |
| RU | Russian Federation |
| SD | Sudan |
| SE | Sweden |
| SG | Singapore |
| SI | Slovenia |
| SK | Slovakia |
| SN | Senegal |
| SZ | Swaziland |
| TD | Chad |
| TG | Togo |
| TJ | Tajikistan |
| TM | Turkmenistan |
| TR | Turkey |
| TT | Trinidad and Tobago |
| UA | Ukraine |
| UG | Uganda |
| US | United States of America |
| UZ | Uzbekistan |
| VN | Viet Nam |
| YU | Yugoslavia |
| ZW | Zimbabwe |

# 32 Human Secreted Proteins

## *Field of the Invention*

This invention relates to newly identified polynucleotides and the polypeptides encoded by these polynucleotides, uses of such polynucleotides and polypeptides, and their production.

## *Background of the Invention*

Unlike bacterium, which exist as a single compartment surrounded by a membrane, human cells and other eucaryotes are subdivided by membranes into many functionally distinct compartments. Each membrane-bounded compartment, or organelle, contains different proteins essential for the function of the organelle. The cell uses "sorting signals," which are amino acid motifs located within the protein, to target proteins to particular cellular organelles.

One type of sorting signal, called a signal sequence, a signal peptide, or a leader sequence, directs a class of proteins to an organelle called the endoplasmic reticulum (ER). The ER separates the membrane-bounded proteins from all other types of proteins. Once localized to the ER, both groups of proteins can be further directed to another organelle called the Golgi apparatus. Here, the Golgi distributes the proteins to vesicles, including secretory vesicles, the cell membrane, lysosomes, and the other organelles.

Proteins targeted to the ER by a signal sequence can be released into the extracellular space as a secreted protein. For example, vesicles containing secreted proteins can fuse with the cell membrane and release their contents into the extracellular space - a process called exocytosis. Exocytosis can occur constitutively or after receipt of a triggering signal. In the latter case, the proteins are stored in secretory vesicles (or secretory granules) until exocytosis is triggered. Similarly, proteins residing on the cell membrane can also be secreted into the extracellular space by proteolytic cleavage of a "linker" holding the protein to the membrane.

Despite the great progress made in recent years, only a small number of genes encoding human secreted proteins have been identified. These secreted proteins include the commercially valuable human insulin, interferon, Factor VIII, human growth hormone, tissue plasminogen activator, and erythropoeitin. Thus, in light of the pervasive role of secreted proteins in human physiology, a need exists for identifying and characterizing novel human secreted proteins and the genes that encode them. This knowledge will allow one to detect, to treat, and to prevent medical disorders by using secreted proteins or the genes that encode them.

## *Summary of the Invention*

The present invention relates to novel polynucleotides and the encoded polypeptides. Moreover, the present invention relates to vectors, host cells, antibodies, and recombinant methods for producing the polypeptides and polynucleotides. Also provided are diagnostic methods for detecting disorders related to the polypeptides, and therapeutic methods for treating such disorders. The invention further relates to screening methods for identifying binding partners of the polypeptides.

## *Detailed Description*

**Definitions**

The following definitions are provided to facilitate understanding of certain terms used throughout this specification.

In the present invention, "isolated" refers to material removed from its original environment (e.g., the natural environment if it is naturally occurring), and thus is altered "by the hand of man" from its natural state. For example, an isolated polynucleotide could be part of a vector or a composition of matter, or could be contained within a cell, and still be "isolated" because that vector, composition of matter, or particular cell is not the original environment of the polynucleotide.

In the present invention, a "secreted" protein refers to those proteins capable of being directed to the ER, secretory vesicles, or the extracellular space as a result of a signal sequence, as well as those proteins released into the extracellular space without necessarily containing a signal sequence. If the secreted protein is released into the extracellular space, the secreted protein can undergo extracellular processing to produce a "mature" protein. Release into the extracellular space can occur by many mechanisms, including exocytosis and proteolytic cleavage.

As used herein , a "polynucleotide" refers to a molecule having a nucleic acid sequence contained in SEQ ID NO:X or the cDNA contained within the clone deposited with the ATCC. For example, the polynucleotide can contain the nucleotide sequence of the full length cDNA sequence, including the 5' and 3' untranslated sequences, the coding region, with or without the signal sequence, the secreted protein coding region, as well as fragments, epitopes, domains, and variants of the nucleic acid sequence. Moreover, as used herein, a "polypeptide" refers to a molecule having the translated amino acid sequence generated from the polynucleotide as broadly defined.

In the present invention, the full length sequence identified as SEQ ID NO:X was often generated by overlapping sequences contained in multiple clones (contig

analysis). A representative clone containing all or most of the sequence for SEQ ID NO:X was deposited with the American Type Culture Collection ("ATCC"). As shown in Table 1, each clone is identified by a cDNA Clone ID (Identifier) and the ATCC Deposit Number. The ATCC is located at 10801 University Boulevard, Manassas, Virginia 20110-2209, USA. The ATCC deposit was made pursuant to the terms of the Budapest Treaty on the international recognition of the deposit of microorganisms for purposes of patent procedure.

A "polynucleotide" of the present invention also includes those polynucleotides capable of hybridizing, under stringent hybridization conditions, to sequences contained in SEQ ID NO:X, the complement thereof, or the cDNA within the clone deposited with the ATCC. "Stringent hybridization conditions" refers to an overnight incubation at 42° C in a solution comprising 50% formamide, 5x SSC (750 mM NaCl, 75 mM sodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 μg/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1x SSC at about 65°C.

Also contemplated are nucleic acid molecules that hybridize to the polynucleotides of the present invention at lower stringency hybridization conditions. Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages of formamide result in lowered stringency); salt conditions, or temperature. For example, lower stringency conditions include an overnight incubation at 37°C in a solution comprising 6X SSPE (20X SSPE = 3M NaCl; 0.2M $NaH_2PO_4$; 0.02M EDTA, pH 7.4), 0.5% SDS, 30% formamide, 100 ug/ml salmon sperm blocking DNA; followed by washes at 50°C with 1XSSPE, 0.1% SDS. In addition, to achieve even lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5X SSC).

Note that variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. Typical blocking reagents include Denhardt's reagent, BLOTTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility.

Of course, a polynucleotide which hybridizes only to polyA+ sequences (such as any 3' terminal polyA+ tract of a cDNA shown in the sequence listing), or to a

complementary stretch of T (or U) residues, would not be included in the definition of "polynucleotide," since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-stranded cDNA clone).

The polynucleotide of the present invention can be composed of any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. For example, polynucleotides can be composed of single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, the polynucleotide can be composed of triple-stranded regions comprising RNA or DNA or both RNA and DNA. A polynucleotide may also contain one or more modified bases or DNA or RNA backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications can be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically, or metabolically modified forms.

The polypeptide of the present invention can be composed of amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres, and may contain amino acids other than the 20 gene-encoded amino acids. The polypeptides may be modified by either natural processes, such as posttranslational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched , for example, as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched, and branched cyclic polypeptides may result from posttranslation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine,

formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, pegylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. (See, for instance, PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993); POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York, pgs. 1-12 (1983); Seifter et al., Meth Enzymol 182:626-646 (1990); Rattan et al., Ann NY Acad Sci 663:48-62 (1992).)

"SEQ ID NO:X" refers to a polynucleotide sequence while "SEQ ID NO:Y" refers to a polypeptide sequence, both sequences identified by an integer specified in Table 1.

"A polypeptide having biological activity" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. In the case where dose dependency does exist, it need not be identical to that of the polypeptide, but rather substantially similar to the dose-dependence in a given activity as compared to the polypeptide of the present invention (i.e., the candidate polypeptide will exhibit greater activity or not more than about 25-fold less and, preferably, not more than about tenfold less activity, and most preferably, not more than about three-fold less activity relative to the polypeptide of the present invention.)

## Polynucleotides and Polypeptides of the Invention

### FEATURES OF PROTEIN ENCODED BY GENE NO: 1

This gene maps to chromosome 3 and therefore polynucleotides of the present invention can be used in linkage analysis as a marker for chromosome 3.

This gene is expressed in several fetal tissues including brain, liver and lung and to a lesser extent in adult tissues, particularly skin.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, a variety of cancers, particularly of the brain, liver, and lung. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a

number of disorders of the above tissues or cells, particularly of the central nervous system, hepatic system, and hepatic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., brain and other tissue of the nervous system, liver, lung, and skin, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful as a target for a variety of blocking agents, as they are likely to be involved in the promotion of a variety of cancers.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 2**

In specific embodiments, the polypeptides of the invention comprise the sequence:MSVPAFIDISEEDQAAELRAYLKSKGAEISEENSEGGLHVDLAQIIEACDVCLKEDDKDVESVMNSVVSLLLILEPDKQEALIESLCEKLVKFREGERPSLRLQLLSNLFHGMDKNTPVRYTVYCSLIKVAASCGAIQYIPTELDQVRKWISDWNLTTEKKHTLLRLLYEALVDCKKSDAASKVMVELLGSYTEDNASQARVDAHRCIVRALKDPNAFLFDHLLTLKPVKFLEGELIHDLLTIFVSAKLASYVKFYQNNKDFIDSLGLLHEQNMAKMRLLTFMGMAVENKEISFDTMQQELQIGADDVEAFVIDAVRTKMVYCKIDQTQRKVVVSHSTHRTFGKQQWQQLYDTLNAWKQNLNKVKNSLLSLSDT (SEQ ID NO:85), MSVPAFIDISEED (SEQ ID NO:86),QAAELRAYLKSKGAE (SEQ ID NO:87), ISEENSEGGLHVDLAQI (SEQ ID NO:88), IEACDVCLKEDDKDVESV (SEQ ID NO:89), VARPSSLFRSAWSCEW (SEQ ID NO:90), LRLQLLSNLFHG (SEQ ID NO:91), KDVESVMNSVVSLLLIL (SEQ ID NO:92), DAASKVMVELLGSYTEDNASQARVDA (SEQ ID NO:93), and/or VEAFVIDAVR (SEQ ID NO:94). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed in bone and to a lesser extent in brain, lung, T-cells, muscle, skin, testis, spleen and macrophages.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, bone cancer, osteoarthritis, and autoimmune diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a

number of disorders of the above tissues or cells, particularly of the immune system and skeletal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, T-cells and other cells and tissue of the immune system, lung, muscle, skin, and testis and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:49 as residues: Arg-31 to Ser-37, Met-50 to Val-56, Glu-80 to Trp-87, Thr-94 to His-99, Tyr-129 to Ser-135, Tyr-193 to Phe-199, Ser-274 to Gln-285, and/or Ala-293 to Lys-302.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 3

The translation product of this gene shares sequence homology with various kinases. The closest homolog is mouse TIF1 which is a mouse nuclear protein. TIF1 enhances RXR and RAR AF-2 in yeast and interacts in a ligand-dependent manner with several nuclear receptors in yeast and mammalian cells, as well as in vitro. Remarkably, these interactions require the amino acids constituting the AF-2 activating domain conserved in all active NRs. Moreover, the oestrogen receptor (ER) AF-2 antagonist hydroxytamoxifen cannot promote ER-TIF1 interaction. We propose that TIF1, which contains several conserved domains found in transcriptional regulatory proteins, is a mediator of ligand-dependent AF-2. Interestingly, the TIF1 N-terminal moiety is fused to B-raf in the mouse oncoprotein T18.

This gene is expressed primarily in activated T-cells and to a lesser extent in various other tissues including testes and brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, autoimmune diseases, AIDS, leukemias, and various other cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other cells and tissue of the immune system, testes and other reproductive tissue, and brain and other

tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:50 as residues: Ala-31 to Glu-36.

The tissue distribution and homology to TIF indicates that polynucleotides and polypeptides corresponding to this gene are useful for modulation of nuclear receptor and ligand interaction in various immune disorders.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 4

This gene maps to chromosome 11. Accordingly, polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 11. In specific embodiments, the polypeptides of the invention comprise the sequence: MSEIYLRCQDEQQYARWMAGCRLASKGRTMADSSY (SEQ ID NO:95), LVAPRF QRKFKAKQLTPRILEAHQNVAQLSLAEAQLRFIQAWQSL (SEQ ID NO:96), VGD VVKTWRFSNMRQWNVNWDIR (SEQ ID NO:97), EEIDCTEEEMMVFAALQYH INKLSQS (SEQ ID NO:98), and/or EEIDCTEEEMMVFAALQYHINKLSQS (SEQ ID NO:99). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in several white blood cell types including monocytes, T-cells, and neutrophils and to a lesser extent in a limited number of other tissues including umbilical vein and liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, various diseases of the immune system including AIDS, immunodeficency diseases, and autoimmune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, liver, and vascular tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily

fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:51 as residues: Ser-3 to Pro-9, Leu-17 to Leu-29, Asp-64 to Pro-69, Ile-105 to Gln-110, Thr-183 to Gln-200, Cys-239 to Arg-247, Ser-256 to Met-261, Gln-280 to Ala-296, Arg-310 to Thr-321, Lys-363 to Asp-368, Ser-395 to Trp-400, and/or Thr-443 to Asp-453.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for replacement therapy in a variety of immune system disorders.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 5

This gene is expressed primarily in brain and little or not at all in any other tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, mood disorders, schizophrenia and related diseases, bipolar disorder and unipolar depression. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:52 as residues: Met-1 to Gly-8, Pro-10 to Arg-17, Pro-45 to Ser-55, and/or Gly-63 to Tyr-74.

The tissue distribution of this gene primarily in brain indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. Also given the brain-specific expression of this gene, the promoter region of this gene contains a brain-specific element that could be used for targeting expression of vector systems to the brain in gene replacement therapy.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 6

This gene maps to chromosome 1 and therefore, polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 1.

This gene is expressed abundantly in rhabdomyosarcoma, is expressed to a high level and in different regions of the brain and pituitary gland and to a lesser extent in a variety of other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neurological disorders and muscular disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., smooth muscle, brain and other tissue of the nervous system, and pituitary, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The abundant expression of this gene in rhadomyosarcoma indicates a role for the protein product either in the detection and/or treatment of skeletal muscle disorders including muscle degeneration, muscle wasting, and rhabdomyolysis. Furthermore expression in the brain indicates a role for the protein product of this gene in the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 7

The translation product of this gene shares sequence homology with the TDAG51 gene which is thought to be important in the mediation of apoptosis and cell death by coupling TCR stimulation to Fas expression. In specific embodiments, the polypeptides of the invention comprise the sequence: KELSFARIKAVECVESTGR HIYFTLV(SEQ ID NO:100) and/or GWNAQITLGLVKFKNQQ (SEQ ID NO:101).

This gene is expressed in various human tissues including macrophages.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., macrophages and other blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:54 as residues: Met-1 to Pro-9, Gln-43 to Glu-49, and/or Phe-95 to Arg-102.

The tissue distribution and homology to TDAG51 gene indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of immune disorders, such as immunodeficiency, allergy, infection, inflammation, tissue/organ transplantation.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 8**

This gene is expressed in breast tissue, and amniotic cells and to a lesser extent in smooth muscle, T-cells, and infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, fetal distress syndrome and embryonic wasting. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the female reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., mammary tissue, amniotic cells, smooth muscle, brain and other tissue of the nervous system, and T-cells and other cells and tissue of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 9

In specific embodiments, the polypeptides of the invention comprise the sequence: LVLGLSXLNNSYNFSF (SEQ ID NO:102), HVVIGSQAEEGQYSLNF (SEQ ID NO:103), HNCNNSVPGKEHPFDITVM (SEQ ID NO:104 ), FIKYVLSD KEKKVFGIV (SEQ ID NO:105), IPMQVLANVAYII (SEQ ID NO:106), IPMQVL ANVAYII (SEQ ID NO:107), DGKVAVNLAKLKLFR (SEQ ID NO:108), and/or IREKNPDGFLSAA (SEQ ID NO:109). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is primarily expressed in the fetal liver, spleen and pituitary gland, and to a lesser extent in multiple tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune disorders and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hepatic, immune and hematopoetic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., liver, spleen, and pituitary gland, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:56 as residues: Ser-62 to Cys-71, Thr-78 to Leu-86, Ser-104 to Lys-109, Ser-130 to Ala-135, and/or Gln-168 to Asp-174.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of hepatic disorders, and disorders of the immune and hematopoetic systems, such as hepatic failure, hepatitis, alcoholic liver diseases, portal hypertension, toxic liver injury, liver transplantation, and neoplasm of the liver. The expression in the fetal liver spleen also indicates its function in hematopoiesis, and therefore the gene may be useful in hematopoietic disorders including anemia, leukemia or cancer

radiotherapy/chemotherapy. The expression in the pituitary gland may indicate its use in endocrine disorders with systemic or specific manifestations.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 10**

The translation product of this gene shares sequence homology with a chicken DNA binding protein which is thought to be important in transcriptional regulation of gene expression. In specific embodiments, polypeptides of the invention comprise the sequence: MMFGGYETI (SEQ ID NO:110), YRDESSSELSVDSEVEFQLYSQIH (SEQ ID NO:111), YAQDLDDVIREEEHEEKNSGNSESSSSKPNQKKLIVLSDSEVI QLSDGSEVITLSDEDSIYRCKGKNVRVQAQENAHGLSSSLQSNELVDKKCKSDI EKPKSEERSGVIREVMIIEVSSSEEEESTISEGDNVESW (SEQ ID NO:112), MLLG CEVDDKDDDILLNLVGCENSVTEGEDGINWSIS (SEQ ID NO:113), DKDIEAQI ANNRTPGRWT (SEQ ID NO:114), QRYYSANKNIICRNCDKRGHLSKNCPLP RKV (SEQ ID NO:115), and/or RRCFLCSRRGHLLYSCPAPLCEYCPVPKMLDHS CLFRHSWDKQCDRCHMLGHYTDACTEIWRQYHLTTKPGPPKKPKTPSRPSAL AYCYHCAQKGHYGHECPEREVYDPSPVSPFICYYXDKYEIQEREKRLKQKIKV XKKNGVIPEPSKLPYIKAANENPHHDIRKGRASWKSNRWPQ (SEQ ID NO:116). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed in tonsils and bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders of the immune, hematopoetic, and lymphatic systems. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, hematopoetic, and lymph systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., tonsils, and bone marrow, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to DNA binding protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of disorders in the immune, hematopoetic, and lymph systems.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 11

This gene is expressed in dendritic and T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders of the immune system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cells types (e.g., dendritic cells, and T-cells and other cells and tissue of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for the treatment and diagnosis of immune system disorders, particularly those involving dendritic or T-cells such as inflammation.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 12

This gene is expressed in activated neutrophils, endothelial cells, T cells and to a lesser extent in brain and liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, AIDS, immune disorders and susceptibility to infectious disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and skin, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., neutrophils and other

blood cells, endothelial cells, T-cells and other cells and tissue of the immune system, brain and other tissue of the nervous system, and liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:59 as residues: Glu-41 to Val-46.

This gene product is useful for the diagnosis and/or treatment of a variety of disorders, including hematopoietic disorders, neurological disorders, liver disease, and disorders involving angiogenesis.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 13

This gene is expressed in keratinocytes and to a lesser extent in endothelial cells and placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, impaired wound healing. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skin, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cells types (e.g., keratinocytes and other cells of the skin, endothelial cells, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:60 as residues: Pro-35 to Trp-42, Ala-53 to Asp-62, and/or Arg-103 to Pro-113.

The tissue distribution indicates that the protein products of this gene are useful for the treatment of wound healing deficiency and skin disorders.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 14

This gene is expressed in kidney and to a lesser extent in embryonic tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions which include, but are not limited to, renal failure. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the kidney, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., kidney, embryonic and other rapidly developing (e.g., dividing) tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 15**

This gene is expressed primarily in brain and to a lesser extent in liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, depression, manic depression and other mental diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., brain and other tissue of the nervous system, and liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for the treatment of central nervous system disorders such as depression and other mental illnesses.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 16**

This gene is expressed in fetal brain and to a lesser extent in placenta, endothelial cells, fetal lung, and T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, restinosis, birth defects and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular system, and developmental process, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, endothelial cells, lung, and T-cells and other cells and tissue of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:63 as residues: Gln-36 to Lys-42, and/or Glu-89 to Arg-104.

The tissue distribution indicates that the protein products of this gene are useful for the development of agonists and/or antagonists for treatment of nervous system disorders and fetal development.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 17

This gene is expressed in hemangiopericytoma and to a lesser extent in fetal tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, hemangiopericytomas and other cancers, as well as developmental disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., vascular tissue, pericytic tissue, and developing tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include

those comprising a sequence shown in SEQ ID NO:64 as residues: Glu-43 to Pro-51, Gly-71 to Arg-82, Pro-96 to Arg-103, and/or Thr-130 to Gly-140.

The polynucleotides and polypeptides related to this gene are believed to be useful for the treatment and diagnosis of tumors, particularly hemangiopericytomas, and for the treatment of developmental disorders.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 18

This gene is expressed in fetal liver and to a lesser extent in brain and T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, fetal disorders, fetal development, and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hepatic system, nervous system and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, brain and other tissue of the nervous system, and T-cells and other cells and tissue of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for the identification of agonists and /or antagonists for treatment of mental illnesses such as schizophrenia and depression. The gene product may also be useful for monitoring fetal development during pregnancy.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 19

This gene is expressed in T cells and to a lesser extent in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, central nervous diseases and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of

disorders of the above tissues or cells, particularly of the central nervous system and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other cells and tissue of the immune system, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:66 as residues: Lys-69 to Leu-74, Ser-92 to Phe-97, Asp-109 to Leu-117, Leu-142 to Ser-159, Thr-166 to Glu-183, Ala-191 to Glu-205, and/or Pro-213 to Glu-220.

The tissue distribution indicates that the protein products of this gene are useful for the development of drugs for treatment of disorders affecting the central nervous system and immune system.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 20

The translation product of this gene shares sequence homology with a *C. elegans* ORF that seems to be a transmembrane protein. (See GenBank Accession No. 790406.) This contig has two probable frameshifts between the +2 and +3 frames based on homology with the *C. elegans* gene. This frameshift can easily be resolved by sequencing the deposited clone. Moreover, this gene maps to chromosome 8, and therefore can be used as a marker in linkage analysis for chromosome 8.

This gene is expressed ubiquitously, including T cells and amygdala.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and endocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other cells and tissue of the immune system, amygdala, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The ubiquitous tissue distribution and homology to a C. elegans transmembrane-like protein indicates that the protein product of this gene plays a role important in both vertebrates and invertebrates and is useful for diagnosis or treatment of disorders related to this gene.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 21

This gene is expressed primarily in embryonic and testes and to a lesser extent in ovary, hepatoma, kidney, endothelial, and smooth muscle cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, metabolic disorder, abnormal embryonic development and tumor. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the embryonic or vascular tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., ovary and other reproductive tissue, kidney, endothelial cells, and smooth muscle cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to NADH dehydrogenase indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and/or treating metabolic disorders, particularly involving embryonic and vascular tissues.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 22

The translation product of this gene shares sequence homology with alpha 1C adrenergic receptor which is thought to be important in neuronal signal transmission.

This gene is expressed primarily in breast lymphnode and to a lesser extent in uterine cancer and testis tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neurological disorders. Similarly, polypeptides and antibodies directed to

these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neurologic, breast lymphonode, uterine cancer, and testis, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., breast tissue, lymphoid tissue, uterine tissue, and testis and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to alpha 1C adrenergic receptor indicates that polynucleotides and polypeptides corresponding to this gene are useful for transmitting signals to neurons.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 23

The translation product of this gene shares sequence homology with G-protein-coupled receptor which is thought to be important in mediating a wide variety of physiological function and belongs to a gene superfamily with members ranging from chemokine receptor to bradykinin receptor. This gene has also recently been cloned by another group, calling the gene platelet activating receptor homolog. (See GenBank Accession No. 2580588.) Preferred polypeptide fragments comprise the amino acid sequence: LSIIFLAFVSIDRCLQL (SEQ ID NO:117) and GSCFATWAFIQKNTNHRCVSIY LINLLTADFLLTLALPVKIVVDLGVAPWKLKIFHCQVTACLIYIN (SEQ ID NO:118). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in immune cells, particularly lymphocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders of lymphocytes and other immune cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., lymphocytes and other cells and tissue of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum,

plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:70 as residues: Asp-59 to Asn-65, Lys-72 to Trp-79, Tyr-110 to Val-121, and/or Ala-204 to Asn-215.

The tissue distribution and homology to G-protein coupled receptor indicates that polynucleotides and polypeptides corresponding to this gene are useful as chemokine receptor on lymphocytes that regulate immune response.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 24

The translation product of this gene shares sequence homology with protein disulfide isomerase which is thought to be important in protein folding and protein-protein interaction. This gene also shares homology to genes having thioredoxin domains. (See Accession No. 1943817.) This gene also maps to chromosome 9, and therefore may be useful in linkage analysis as a marker for chromosome 9.

This gene is expressed primarily in tumor tissues and to a lesser extent in a wide variety of normal tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders due to inappropriate protein folding and protein-protein interaction. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the tumorigenic process, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:71 as residues: Glu-78 to Asn-83, Asp-91 to Gln-100, Glu-122 to Ser-128, Arg-137 to Pro-143, Asp-157 to Asn-162, Glu-168 to Asn-174, Ser-199 to Gly-206, Pro-213 to Ala-218, Glu-251 to Thr-257, Ser-353 to His-361, Gly-363 to Ala-375, Pro-382 to Phe-387, and/or Arg-401 to Leu-406.

The tissue distribution and homology to protein disulfide isomerase indicates that polynucleotides and polypeptides corresponding to this gene are useful for regulating protein folding and protein-protein interaction in tumor tissues.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 25**

This gene is expressed primarily in leukocytes involved in immune defense, including T cells, macrophages, neutrophils and to a lesser extent in synovium, adrenal gland tumor, adipose, and placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, defects or disorders in leukocytes. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and defense systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., leukocytes and other cells and tissues of the immune system, synovium, adrenal gland, adipose and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for regulating leukocyte function and may be used for diagnosis and treatment of disorders in immune and defense systems.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 26**

This gene is expressed in a variety of tissues and cell types, including colon cancer, breast cancer, neutrophils, T cells, spinal cord, fibroblasts, and vascular endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer, disorder and abnormalities in leukocytes and other tissues. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell

type(s). For a number of disorders of the above tissues or cells, particularly those cells involved in tumorigenesis and immune defense systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., colon, breast tissue, neutrophils, T-cells and other blood cells, spinal cord and other tissue of the nervous system, endothelial cells, vascular tissue, and fibroblasts, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cancer or immune system disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 27**

The translation product of this gene shares sequence homology with a mouse pancreatic polypeptide. (See GenBankAccession No. 200464.) Thus, it is likely that this gene has activity similar to the mouse pancreatic polypeptide. Preferred polypeptide fragments comprise the amino acids sequence: APLETMQNKPRAPQKRALPFPEL ELRDYASVLTRYSLGLRNKEPSLGHRWGTQKLGRSPC (SEQ ID NO:119). Also preferred are polynucleotide fragments encoding this polypeptide fragment.

This gene is expressed primarily in neutrophils and to a lesser extent in induced endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders in neutrophils or leukocyte adhesion. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., neutrophils and other blood cells, and endothelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for regulation of neutrophils or leukocyte adhesion to endothelial cells. It may be used to diagnose or treat disorders associated with neutrophils and vascular endothelial cells.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 28

This gene is expressed primarily in prostate BPH.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, benign hypertrophy of the prostate. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male urogenital system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., prostate, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of benign hypertrophy of the prostate or prostate cancer.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 29

The translation product of this gene shares sequence homology with C16C10.7, a *C. elegans* gene similar to zinc finger protein, a protein involved in DNA binding. Thus, this protein is expected to share certain biological activities with C16C10.7 including DNA binding activities.

This gene is expressed primarily in activated T-cells and to a lesser extent in fetal brain, TNF-induced amniotic cells and epididymus.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune or neurodegenerative disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes

for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and central nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other cells and tissue of the immune system, brain and other tissue of the nervous system, amniotic cells, and epididymus and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for the diagnosis and treatment of immune and/or neurodegenerative disorders and promotion of survival and differentiation of neurons.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 30**

This gene is expressed primarily in T-cells and to a lesser extent in bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immunological disorders including autoimmune disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other cells and tissue of the immune system, and bone marrow, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. It is believed that this gene maps to chromosome 4: Transcript map: WI-11395, Chr.4, D4S395-D4S414; Whitehead map: WI-11395, Chr.4, 498.0 cR; dbSTS entries: G21269.

The tissue distribution indicates that the protein products of this gene are useful for diagnosis and treatment of immunologically mediated disorders as they are thought

to play a role in the proliferation, survival, differentiation, and/or activation of a variety of hematopoietic cells, including early progenitors or hematopoietic stem cells.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 31

This gene is expressed primarily in human skin.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, wound healing and skin cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the integumentary system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., skin and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for diagnosis and treatment of skin cancers and wound healing.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 32

The translation product of this gene shares sequence homology with human Tear Prealbumin (GenBank accession no. gi|307518) and rat Oderant-binding protein (GenBank accession no. gi|207551), both of which are thought to be important in molecule binding and transport.

This gene is expressed primarily in endometrial tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancers of the endometrium, skin and haemopoietic system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the haemopoietic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cells and tissue of the immune system, and

endometrium and other tissue of the reproductive system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to the molecule binding and transport gene family indicates that the protein products of this gene are useful for the diagnosis and treatment of cancers of the endometrium and haemopoietic system as well as for the treatment of autoimmune disorders such as inflammation.

| Gene No. | cDNA Clone ID | ATCC Deposit Nr and Date | Vector | NT SEQ ID NO: X | Total NT Seq. | 5' NT of Clone Seq. | 3' NT of Clone Seq. | 5' NT of Start Codon | 5' NT of First AA of Signal Pep | AA SEQ ID NO: Y | First AA of Sig Pep | Last AA of Sig Pep | First AA of Secreted Portion | Last AA of ORF |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | HSVBZ80 | 209075 05/22/97 | Uni-ZAP XR | 11 | 1169 | 64 | 1060 | 162 | 162 | 48 | 1 | 38 | 39 | 145 |
| 2 | HTAAU21 | 209075 05/22/97 | Uni-ZAP XR | 12 | 1310 | 1 | 1310 | 283 | 283 | 49 | 1 | 18 | 19 | 311 |
| 3 | HTLEK16 | 209075 05/22/97 | Uni-ZAP XR | 13 | 1139 | 19 | 1111 | | 251 | 50 | 1 | 21 | 22 | 46 |
| 4 | HUSIR91 | 209075 05/22/97 | pSport1 | 14 | 2271 | 743 | 2271 | 59 | 59 | 51 | 1 | 23 | 24 | 466 |
| 4 | HUSIR91 | 209075 05/22/97 | pSport1 | 43 | 2581 | 1035 | 2164 | 1148 | 1148 | 80 | 1 | 27 | 28 | 207 |
| 5 | HADMC21 | 209075 05/22/97 | pBluescript | 15 | 626 | 60 | 479 | 91 | 91 | 52 | 1 | 51 | 52 | 82 |
| 6 | HAGFM45 | 209075 05/22/97 | Uni-ZAP XR | 16 | 2118 | 1170 | 2058 | 1248 | 1248 | 53 | 1 | 16 | 17 | 62 |
| 7 | HAIBE65 | 209075 05/22/97 | Uni-ZAP XR | 17 | 1076 | 396 | 993 | 528 | 528 | 54 | 1 | 31 | 32 | 123 |
| 8 | HAQBH57 | 209075 05/22/97 | Uni-ZAP XR | 18 | 1379 | 420 | 1306 | 618 | 618 | 55 | 1 | 25 | 26 | 179 |
| 9 | HATCX80 | 209075 05/22/97 | Uni-ZAP XR | 19 | 1337 | 47 | 1337 | 199 | 199 | 56 | 1 | 18 | 19 | 286 |
| 10 | HCFLQ84 | 209075 05/22/97 | pSport1 | 20 | 1390 | 237 | 1390 | 410 | 410 | 57 | 1 | 20 | 21 | 33 |
| 11 | HCFLS78 | 209075 05/22/97 | pSport1 | 21 | 1431 | 178 | 981 | 420 | 420 | 58 | 1 | 21 | 22 | 23 |

| Gene No. | cDNA Clone ID | ATCC Deposit Nr and Date | Vector | NT SEQ ID NO: X | Total NT Seq. | 5' NT of Clone Seq. | 3' NT of Clone Seq. | 5' NT of Start Codon | 5' NT of First AA of Signal Pep | AA SEQ ID NO: Y | First AA of Sig Pep | Last AA of Sig Pep | First AA of Secreted Portion | Last AA of ORF |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 12 | HTADI12 | 209075 05/22/97 | Uni-ZAP XR | 22 | 2539 | 69 | 2539 | 104 | 104 | 59 | 1 | 27 | 28 | 46 |
| 13 | HEMCM42 | 209075 05/22/97 | Uni-ZAP XR | 23 | 1041 | 48 | 1007 | 58 | 58 | 60 | 1 | 29 | 30 | 113 |
| 14 | HEONP72 | 209075 05/22/97 | pSport1 | 24 | 1962 | 1 | 1947 | 181 | 181 | 61 | 1 | 19 | 20 | 31 |
| 15 | HFCDW34 | 209075 05/22/97 | Uni-ZAP XR | 25 | 1228 | 321 | 1228 | 525 | 525 | 62 | 1 | 24 | 25 | 80 |
| 16 | HTTEU91 | 209075 05/22/97 | Uni-ZAP XR | 26 | 1340 | 325 | 1340 | 15 | 15 | 63 | 1 | 18 | 19 | 103 |
| 17 | HHGBF89 | 209075 05/22/97 | Lambda ZAP II | 27 | 806 | 31 | 806 | 77 | 77 | 64 | 1 | 19 | 20 | 145 |
| 17 | HHGBF89 | 209075 05/22/97 | Lambda ZAP II | 45 | 796 | 31 | 796 | 77 | 77 | 82 | 1 | 25 | 26 | 145 |
| 18 | HKIYQ65 | 209075 05/22/97 | pBluescript | 28 | 696 | 1 | 684 | 98 | 98 | 65 | 1 | 17 | 18 | 30 |
| 19 | HKMLN27 | 209075 05/22/97 | pBluescript | 29 | 1007 | 71 | 963 | 129 | 129 | 66 | 1 | 23 | 24 | 259 |
| 20 | HKIAC30 | 209022 05/08/97 | Uni-ZAP XR | 30 | 2017 | 126 | 2007 | 161 | 161 | 67 | 1 | | | 22 |
| 21 | HKIXB95 | 209022 05/08/97 | pBluescript | 31 | 699 | 196 | 699 | 230 | 230 | 68 | 1 | 22 | 23 | 26 |
| 22 | HLMIY86 | 209022 05/08/97 | Lambda ZAP II | 32 | 1264 | 1 | 1264 | | 342 | 69 | 1 | 16 | 17 | 28 |
| 23 | HLYAZ61 | 209022 05/08/97 | pSport1 | 33 | 997 | 74 | 997 | 205 | 205 | 70 | 1 | 20 | 21 | 215 |

| Gene No. | cDNA Clone ID | ATCC Deposit Nr and Date | Vector | NT SEQ ID NO: X | Total NT Seq. | 5' NT of Clone Seq. | 3' NT of Clone Seq. | 5' NT of Start Codon | 5' NT of First AA of Signal Pep | AA SEQ ID NO: Y | First AA of Sig Pep | Last AA of Sig Pep | First AA of Secreted Portion | Last AA of ORF |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 24 | HMQDT36 | 209022 05/08/97 | Uni-ZAP XR | 34 | 1914 | 37 | 1897 | 192 | 192 | 71 | 1 | 32 | 33 | 406 |
| 25 | HNEDF25 | 209022 05/08/97 | Uni-ZAP XR | 35 | 1020 | 11 | 1020 | | 211 | 72 | 1 | | | 8 |
| 26 | HNFET17 | 209022 05/08/97 | Uni-ZAP XR | 36 | 781 | 31 | 781 | 100 | 100 | 73 | 1 | | | 33 |
| 27 | HNHCR46 | 209022 05/08/97 | Uni-ZAP XR | 37 | 966 | 507 | 948 | | 576 | 74 | 1 | 29 | 30 | 56 |
| 28 | HPWAS91 | 209022 05/08/97 | Uni-ZAP XR | 38 | 416 | 1 | 416 | 95 | 95 | 75 | 1 | 24 | 25 | 25 |
| 29 | HWTAW41 | 209022 05/08/97 | Uni-ZAP XR | 39 | 1114 | 804 | 1114 | 843 | 843 | 76 | 1 | | | 14 |
| 30 | HBMUT52 | 209022 05/08/97 | Uni-ZAP XR | 40 | 602 | 142 | 602 | 204 | 204 | 77 | 1 | 26 | 27 | 32 |
| 31 | HERAG83 | 209022 05/08/97 | Uni-ZAP XR | 41 | 970 | 1 | 970 | 110 | 110 | 78 | 1 | 22 | 23 | 22 |
| 32 | HETFI51 | 209022 05/08/97 | Uni-ZAP XR | 42 | 1002 | 1 | 1002 | 43 | 43 | 79 | 1 | 21 | 22 | 172 |
| 32 | HETFI51 | 209022 05/08/97 | Uni-ZAP XR | 47 | 981 | 1 | 981 | 23 | 23 | 84 | 1 | 17 | 18 | 30 |

Table 1 summarizes the information corresponding to each "Gene No." described above. The nucleotide sequence identified as "NT SEQ ID NO:X" was assembled from partially homologous ("overlapping") sequences obtained from the "cDNA clone ID" identified in Table 1 and, in some cases, from additional related DNA clones. The overlapping sequences were assembled into a single contiguous sequence of high redundancy (usually three to five overlapping sequences at each nucleotide position), resulting in a final sequence identified as SEQ ID NO:X.

The cDNA Clone ID was deposited on the date and given the corresponding deposit number listed in "ATCC Deposit No:Z and Date." Some of the deposits contain multiple different clones corresponding to the same gene. "Vector" refers to the type of vector contained in the cDNA Clone ID.

"Total NT Seq." refers to the total number of nucleotides in the contig identified by "Gene No." The deposited clone may contain all or most of these sequences, reflected by the nucleotide position indicated as "5' NT of Clone Seq." and the "3' NT of Clone Seq." of SEQ ID NO:X. The nucleotide position of SEQ ID NO:X of the putative start codon (methionine) is identified as "5' NT of Start Codon." Similarly , the nucleotide position of SEQ ID NO:X of the predicted signal sequence is identified as "5' NT of First AA of Signal Pep."

The translated amino acid sequence, beginning with the methionine, is identified as "AA SEQ ID NO:Y," although other reading frames can also be easily translated using known molecular biology techniques. The polypeptides produced by these alternative open reading frames are specifically contemplated by the present invention.

The first and last amino acid position of SEQ ID NO:Y of the predicted signal peptide is identified as "First AA of Sig Pep" and "Last AA of Sig Pep." The predicted first amino acid position of SEQ ID NO:Y of the secreted portion is identified as "Predicted First AA of Secreted Portion." Finally, the amino acid position of SEQ ID NO:Y of the last amino acid in the open reading frame is identified as "Last AA of ORF."

SEQ ID NO:X and the translated SEQ ID NO:Y are sufficiently accurate and otherwise suitable for a variety of uses well known in the art and described further below. For instance, SEQ ID NO:X is useful for designing nucleic acid hybridization probes that will detect nucleic acid sequences contained in SEQ ID NO:X or the cDNA contained in the deposited clone. These probes will also hybridize to nucleic acid molecules in biological samples, thereby enabling a variety of forensic and diagnostic methods of the invention. Similarly, polypeptides identified from SEQ ID NO:Y may

be used to generate antibodies which bind specifically to the secreted proteins encoded by the cDNA clones identified in Table 1.

Nevertheless, DNA sequences generated by sequencing reactions can contain sequencing errors. The errors exist as misidentified nucleotides, or as insertions or deletions of nucleotides in the generated DNA sequence. The erroneously inserted or deleted nucleotides cause frame shifts in the reading frames of the predicted amino acid sequence. In these cases, the predicted amino acid sequence diverges from the actual amino acid sequence, even though the generated DNA sequence may be greater than 99.9% identical to the actual DNA sequence (for example, one base insertion or deletion in an open reading frame of over 1000 bases).

Accordingly, for those applications requiring precision in the nucleotide sequence or the amino acid sequence, the present invention provides not only the generated nucleotide sequence identified as SEQ ID NO:X and the predicted translated amino acid sequence identified as SEQ ID NO:Y, but also a sample of plasmid DNA containing a human cDNA of the invention deposited with the ATCC, as set forth in Table 1. The nucleotide sequence of each deposited clone can readily be determined by sequencing the deposited clone in accordance with known methods. The predicted amino acid sequence can then be verified from such deposits. Moreover, the amino acid sequence of the protein encoded by a particular clone can also be directly determined by peptide sequencing or by expressing the protein in a suitable host cell containing the deposited human cDNA, collecting the protein, and determining its sequence.

The present invention also relates to the genes corresponding to SEQ ID NO:X, SEQ ID NO:Y, or the deposited clone. The corresponding gene can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include preparing probes or primers from the disclosed sequence and identifying or amplifying the corresponding gene from appropriate sources of genomic material.

Also provided in the present invention are species homologs. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source for the desired homologue.

The polypeptides of the invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

The polypeptides may be in the form of the secreted protein, including the mature form, or may be a part of a larger protein, such as a fusion protein (see below). It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification , such as multiple histidine residues, or an additional sequence for stability during recombinant production.

The polypeptides of the present invention are preferably provided in an isolated form, and preferably are substantially purified. A recombinantly produced version of a polypeptide, including the secreted polypeptide, can be substantially purified by the one-step method described in Smith and Johnson, Gene 67:31-40 (1988). Polypeptides of the invention also can be purified from natural or recombinant sources using antibodies of the invention raised against the secreted protein in methods which are well known in the art.

**Signal Sequences**

Methods for predicting whether a protein has a signal sequence, as well as the cleavage point for that sequence, are available. For instance, the method of McGeoch, Virus Res. 3:271-286 (1985), uses the information from a short N-terminal charged region and a subsequent uncharged region of the complete (uncleaved) protein. The method of von Heinje, Nucleic Acids Res. 14:4683-4690 (1986) uses the information from the residues surrounding the cleavage site, typically residues -13 to +2, where +1 indicates the amino terminus of the secreted protein. The accuracy of predicting the cleavage points of known mammalian secretory proteins for each of these methods is in the range of 75-80%. (von Heinje, supra.) However, the two methods do not always produce the same predicted cleavage point(s) for a given protein.

In the present case, the deduced amino acid sequence of the secreted polypeptide was analyzed by a computer program called SignalP (Henrik Nielsen et al., Protein Engineering 10:1-6 (1997)), which predicts the cellular location of a protein based on the amino acid sequence. As part of this computational prediction of localization, the methods of McGeoch and von Heinje are incorporated. The analysis of the amino acid sequences of the secreted proteins described herein by this program provided the results shown in Table 1.

As one of ordinary skill would appreciate, however, cleavage sites sometimes vary from organism to organism and cannot be predicted with absolute certainty. Accordingly, the present invention provides secreted polypeptides having a sequence shown in SEQ ID NO:Y which have an N-terminus beginning within 5 residues (i.e., + or - 5 residues) of the predicted cleavage point. Similarly, it is also recognized that in

some cases, cleavage of the signal sequence from a secreted protein is not entirely uniform, resulting in more than one secreted species. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

Moreover, the signal sequence identified by the above analysis may not necessarily predict the naturally occurring signal sequence. For example, the naturally occurring signal sequence may be further upstream from the predicted signal sequence. However, it is likely that the predicted signal sequence will be capable of directing the secreted protein to the ER. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

## Polynucleotide and Polypeptide Variants

"Variant" refers to a polynucleotide or polypeptide differing from the polynucleotide or polypeptide of the present invention, but retaining essential properties thereof. Generally, variants are overall closely similar, and, in many regions, identical to the polynucleotide or polypeptide of the present invention.

By a polynucleotide having a nucleotide sequence at least, for example, 95% "identical" to a reference nucleotide sequence of the present invention, it is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence encoding the polypeptide. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. The query sequence may be an entire sequence shown inTable 1, the ORF (open reading frame), or any fragement specified as described herein.

As a practical matter, whether any particular nucleic acid molecule or polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to a nucleotide sequence of the presence invention can be determined conventionally using known computer programs. A preferred method for determing the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and subject sequences are both DNA sequences. An RNA sequence can be compared by converting U's to T's. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB alignment of DNA sequences to calculate percent identiy are:

Matrix=Unitary, k-tuple=4, Mismatch Penalty=1, Joining Penalty=30, Randomization Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty 0.05, Window Size=500 or the lenght of the subject nucleotide sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence because of 5' or 3' deletions, not because of internal deletions, a manual correction must be made to the results. This is becuase the FASTDB program does not account for 5' and 3' truncations of the subject sequence when calculating percent identity. For subject sequences truncated at the 5' or 3' ends, relative to the the query sequence, the percent identity is corrected by calculating the number of bases of the query sequence that are 5' and 3' of the subject sequence, which are not matched/aligned, as a percent of the total bases of the query sequence. Whether a nucleotide is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This corrected score is what is used for the purposes of the present invention. Only bases outside the 5' and 3' bases of the subject sequence, as displayed by the FASTDB alignment, which are not matched/aligned with the query sequence, are calculated for the purposes of manually adjusting the percent identity score.

For example, a 90 base subject sequence is aligned to a 100 base query sequence to determine percent identity. The deletions occur at the 5' end of the subject sequence and therefore, the FASTDB alignment does not show a matched/alignement of the first 10 bases at 5' end. The 10 unpaired bases represent 10% of the sequence (number of bases at the 5' and 3' ends not matched/total number of bases in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 bases were perfectly matched the final percent identity would be 90%. In another example, a 90 base subject sequence is compared with a 100 base query sequence. This time the deletions are internal deletions so that there are no bases on the 5' or 3' of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only bases 5' and 3' of the subject sequence which are not matched/aligned with the query sequnce are manually corrected for. No other manual corrections are to made for the purposes of the present invention.

By a polypeptide having an amino acid sequence at least, for example, 95% "identical" to a query amino acid sequence of the present invention, it is intended that the amino acid sequence of the subject polypeptide is identical to the query sequence except that the subject polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the query amino acid sequence. In other words,

to obtain a polypeptide having an amino acid sequence at least 95% identical to a query amino acid sequence, up to 5% of the amino acid residues in the subject sequence may be inserted, deleted, (indels) or substituted with another amino acid. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

As a practical matter, whether any particular polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to, for instance, the amino acid sequences shown in Table 1 or to the amino acid sequence encoded by deposited DNA clone can be determined conventionally using known computer programs. A preferred method for determing the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and subject sequences are either both nucleotide sequences or both amino acid sequences. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB amino acid alignment are: Matrix=PAM 0, k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Window Size=sequence length, Gap Penalty=5, Gap Size Penalty=0.05, Window Size=500 or the length of the subject amino acid sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence due to N- or C-terminal deletions, not because of internal deletions, a manual correction must be made to the results. This is bécuase the FASTDB program does not account for N- and C-terminal truncations of the subject sequence when calculating global percent identity. For subject sequences truncated at the N- and C-termini, relative to the the query sequence, the percent identity is corrected by calculating the number of residues of the query sequence that are N- and C-terminal of the subject sequence, which are not matched/aligned with a corresponding subject residue, as a percent of the total bases of the query sequence. Whether a residue is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This final percent identity score is what is used for the purposes of the present invention. Only residues to the N- and C-termini of the subject sequence, which are not matched/aligned with the query sequence, are considered for the purposes of manually adjusting the percent identity score. That is, only query residue positions outside the farthest N- and C-terminal residues of the

subject sequence.

For example, a 90 amino acid residue subject sequence is aligned with a 100 residue query sequence to determine percent identity. The deletion occurs at the N-terminus of the subject sequence and therefore, the FASTDB alignment does not show a matching/alignment of the first 10 residues at the N-terminus. The 10 unpaired residues represent 10% of the sequence (number of residues at the N- and C- termini not matched/total number of residues in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 residues were perfectly matched the final percent identity would be 90%. In another example, a 90 residue subject sequence is compared with a 100 residue query sequence. This time the deletions are internal deletions so there are no residues at the N- or C-termini of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only residue positions outside the N- and C-terminal ends of the subject sequence, as displayed in the FASTDB alignment, which are not matched/aligned with the query sequnce are manually corrected for. No other manual corrections are to made for the purposes of the present invention.

The variants may contain alterations in the coding regions, non-coding regions, or both. Especially preferred are polynucleotide variants containing alterations which produce silent substitutions, additions, or deletions, but do not alter the properties or activities of the encoded polypeptide. Nucleotide variants produced by silent substitutions due to the degeneracy of the genetic code are preferred. Moreover, variants in which 5-10, 1-5, or 1-2 amino acids are substituted, deleted, or added in any combination are also preferred. Polynucleotide variants can be produced for a variety of reasons, e.g., to optimize codon expression for a particular host (change codons in the human mRNA to those preferred by a bacterial host such as E. coli).

Naturally occurring variants are called "allelic variants," and refer to one of several alternate forms of a gene occupying a given locus on a chromosome of an organism. (Genes II, Lewin, B., ed., John Wiley & Sons, New York (1985).) These allelic variants can vary at either the polynucleotide and/or polypeptide level. Alternatively, non-naturally occurring variants may be produced by mutagenesis techniques or by direct synthesis.

Using known methods of protein engineering and recombinant DNA technology, variants may be generated to improve or alter the characteristics of the polypeptides of the present invention. For instance, one or more amino acids can be deleted from the N-terminus or C-terminus of the secreted protein without substantial loss of biological function. The authors of Ron et al., J. Biol. Chem. 268: 2984-2988

(1993), reported variant KGF proteins having heparin binding activity even after deleting 3, 8, or 27 amino-terminal amino acid residues. Similarly, Interferon gamma exhibited up to ten times higher activity after deleting 8-10 amino acid residues from the carboxy terminus of this protein. (Dobeli et al., J. Biotechnology 7:199-216 (1988).)

Moreover, ample evidence demonstrates that variants often retain a biological activity similar to that of the naturally occurring protein. For example, Gayle and coworkers (J. Biol. Chem 268:22105-22111 (1993)) conducted extensive mutational analysis of human cytokine IL-1a. They used random mutagenesis to generate over 3,500 individual IL-1a mutants that averaged 2.5 amino acid changes per variant over the entire length of the molecule. Multiple mutations were examined at every possible amino acid position. The investigators found that "[m]ost of the molecule could be altered with little effect on either [binding or biological activity]." (See, Abstract.) In fact, only 23 unique amino acid sequences, out of more than 3,500 nucleotide sequences examined, produced a protein that significantly differed in activity from wild-type.

Furthermore, even if deleting one or more amino acids from the N-terminus or C-terminus of a polypeptide results in modification or loss of one or more biological functions, other biological activities may still be retained. For example, the ability of a deletion variant to induce and/or to bind antibodies which recognize the secreted form will likely be retained when less than the majority of the residues of the secreted form are removed from the N-terminus or C-terminus. Whether a particular polypeptide lacking N- or C-terminal residues of a protein retains such immunogenic activities can readily be determined by routine methods described herein and otherwise known in the art.

Thus, the invention further includes polypeptide variants which show substantial biological activity. Such variants include deletions, insertions, inversions, repeats, and substitutions selected according to general rules known in the art so as have little effect on activity. For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided in Bowie, J. U. et al., Science 247:1306-1310 (1990), wherein the authors indicate that there are two main strategies for studying the tolerance of an amino acid sequence to change.

The first strategy exploits the tolerance of amino acid substitutions by natural selection during the process of evolution. By comparing amino acid sequences in different species, conserved amino acids can be identified. These conserved amino acids are likely important for protein function. In contrast, the amino acid positions where substitutions have been tolerated by natural selection indicates that these

positions are not critical for protein function. Thus, positions tolerating amino acid substitution could be modified while still maintaining biological activity of the protein.

The second strategy uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene to identify regions critical for protein function. For example, site directed mutagenesis or alanine-scanning mutagenesis (introduction of single alanine mutations at every residue in the molecule) can be used. (Cunningham and Wells, Science 244:1081-1085 (1989).) The resulting mutant molecules can then be tested for biological activity.

As the authors state, these two strategies have revealed that proteins are surprisingly tolerant of amino acid substitutions. The authors further indicate which amino acid changes are likely to be permissive at certain amino acid positions in the protein. For example, most buried (within the tertiary structure of the protein) amino acid residues require nonpolar side chains, whereas few features of surface side chains are generally conserved. Moreover, tolerated conservative amino acid substitutions involve replacement of the aliphatic or hydrophobic amino acids Ala, Val, Leu and Ile; replacement of the hydroxyl residues Ser and Thr; replacement of the acidic residues Asp and Glu; replacement of the amide residues Asn and Gln, replacement of the basic residues Lys, Arg, and His; replacement of the aromatic residues Phe, Tyr, and Trp, and replacement of the small-sized amino acids Ala, Ser, Thr, Met, and Gly.

Besides conservative amino acid substitution, variants of the present invention include (i) substitutions with one or more of the non-conserved amino acid residues, where the substituted amino acid residues may or may not be one encoded by the genetic code, or (ii) substitution with one or more of amino acid residues having a substituent group, or (iii) fusion of the mature polypeptide with another compound, such as a compound to increase the stability and/or solubility of the polypeptide (for example, polyethylene glycol), or (iv) fusion of the polypeptide with additional amino acids, such as an IgG Fc fusion region peptide, or leader or secretory sequence, or a sequence facilitating purification. Such variant polypeptides are deemed to be within the scope of those skilled in the art from the teachings herein.

For example, polypeptide variants containing amino acid substitutions of charged amino acids with other charged or neutral amino acids may produce proteins with improved characteristics, such as less aggregation. Aggregation of pharmaceutical formulations both reduces activity and increases clearance due to the aggregate's immunogenic activity. (Pinckard et al., Clin. Exp. Immunol. 2:331-340 (1967); Robbins et al., Diabetes 36: 838-845 (1987); Cleland et al., Crit. Rev. Therapeutic Drug Carrier Systems 10:307-377 (1993).)

**Polynucleotide and Polypeptide Fragments**

In the present invention, a "polynucleotide fragment" refers to a short polynucleotide having a nucleic acid sequence contained in the deposited clone or shown in SEQ ID NO:X. The short nucleotide fragments are preferably at least about 15 nt, and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably, at least about 40 nt in length. A fragment "at least 20 nt in length," for example, is intended to include 20 or more contiguous bases from the cDNA sequence contained in the deposited clone or the nucleotide sequence shown in SEQ ID NO:X. These nucleotide fragments are useful as diagnostic probes and primers as discussed herein. Of course, larger fragments (e.g., 50, 150, 500, 600, 2000 nucleotides) are preferred.

Moreover, representative examples of polynucleotide fragments of the invention, include, for example, fragments having a sequence from about nucleotide number 1-50, 51-100, 101-150, 151-200, 201-250, 251-300, 301-350, 351-400, 401-450, 451-500, 501-550, 551-600, 651-700, 701-750, 751-800, 800-850, 851-900, 901-950, 951-1000, 1001-1050, 1051-1100, 1101-1150, 1151-1200, 1201-1250, 1251-1300, 1301-1350, 1351-1400, 1401-1450, 1451-1500, 1501-1550, 1551-1600, 1601-1650, 1651-1700, 1701-1750, 1751-1800, 1801-1850, 1851-1900, 1901-1950, 1951-2000, or 2001 to the end of SEQ ID NO:X or the cDNA contained in the deposited clone. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) nucleotides, at either terminus or at both termini. Preferably, these fragments encode a polypeptide which has biological activity. More preferably, these polynucleotides can be used as probes or primers as discussed herein.

In the present invention, a "polypeptide fragment" refers to a short amino acid sequence contained in SEQ ID NO:Y or encoded by the cDNA contained in the deposited clone. Protein fragments may be "free-standing," or comprised within a larger polypeptide of which the fragment forms a part or region, most preferably as a single continuous region. Representative examples of polypeptide fragments of the invention, include, for example, fragments from about amino acid number 1-20, 21-40, 41-60, 61-80, 81-100, 102-120, 121-140, 141-160, or 161 to the end of the coding region. Moreover, polypeptide fragments can be about 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, or 150 amino acids in length. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) amino acids, at either extreme or at both extremes.

Preferred polypeptide fragments include the secreted protein as well as the mature form. Further preferred polypeptide fragments include the secreted protein or

the mature form having a continuous series of deleted residues from the amino or the carboxy terminus, or both. For example, any number of amino acids, ranging from 1-60, can be deleted from the amino terminus of either the secreted polypeptide or the mature form. Similarly, any number of amino acids, ranging from 1-30, can be deleted from the carboxy terminus of the secreted protein or mature form. Furthermore, any combination of the above amino and carboxy terminus deletions are preferred. Similarly, polynucleotide fragments encoding these polypeptide fragments are also preferred.

Also preferred are polypeptide and polynucleotide fragments characterized by structural or functional domains, such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions. Polypeptide fragments of SEQ ID NO:Y falling within conserved domains are specifically contemplated by the present invention. Moreover, polynucleotide fragments encoding these domains are also contemplated.

Other preferred fragments are biologically active fragments. Biologically active fragments are those exhibiting activity similar, but not necessarily identical, to an activity of the polypeptide of the present invention. The biological activity of the fragments may include an improved desired activity, or a decreased undesirable activity.

__Epitopes & Antibodies__

In the present invention, "epitopes" refer to polypeptide fragments having antigenic or immunogenic activity in an animal, especially in a human. A preferred embodiment of the present invention relates to a polypeptide fragment comprising an epitope, as well as the polynucleotide encoding this fragment. A region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope." In contrast, an "immunogenic epitope" is defined as a part of a protein that elicits an antibody response. (See, for instance, Geysen et al., Proc. Natl. Acad. Sci. USA 81:3998- 4002 (1983).)

Fragments which function as epitopes may be produced by any conventional means. (See, e.g., Houghten, R. A., Proc. Natl. Acad. Sci. USA 82:5131-5135 (1985) further described in U.S. Patent No. 4,631,211.)

In the present invention, antigenic epitopes preferably contain a sequence of at least seven, more preferably at least nine, and most preferably between about 15 to about 30 amino acids. Antigenic epitopes are useful to raise antibodies, including

monoclonal antibodies, that specifically bind the epitope. (See, for instance, Wilson et al., Cell 37:767-778 (1984); Sutcliffe, J. G. et al., Science 219:660-666 (1983).)

Similarly, immunogenic epitopes can be used to induce antibodies according to methods well known in the art. (See, for instance, Sutcliffe et al., supra; Wilson et al., supra; Chow, M. et al., Proc. Natl. Acad. Sci. USA 82:910-914; and Bittle, F. J. et al., J. Gen. Virol. 66:2347-2354 (1985).) A preferred immunogenic epitope includes the secreted protein. The immunogenic epitopes may be presented together with a carrier protein, such as an albumin, to an animal system (such as rabbit or mouse) or, if it is long enough (at least about 25 amino acids), without a carrier. However, immunogenic epitopes comprising as few as 8 to 10 amino acids have been shown to be sufficient to raise antibodies capable of binding to, at the very least, linear epitopes in a denatured polypeptide (e.g., in Western blotting.)

As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as antibody fragments (such as, for example, Fab and F(ab')2 fragments) which are capable of specifically binding to protein. Fab and F(ab')2 fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding than an intact antibody. (Wahl et al., J. Nucl. Med. 24:316-325 (1983).) Thus, these fragments are preferred, as well as the products of a FAB or other immunoglobulin expression library. Moreover, antibodies of the present invention include chimeric, single chain, and humanized antibodies.

## Fusion Proteins

Any polypeptide of the present invention can be used to generate fusion proteins. For example, the polypeptide of the present invention, when fused to a second protein, can be used as an antigenic tag. Antibodies raised against the polypeptide of the present invention can be used to indirectly detect the second protein by binding to the polypeptide. Moreover, because secreted proteins target cellular locations based on trafficking signals, the polypeptides of the present invention can be used as targeting molecules once fused to other proteins.

Examples of domains that can be fused to polypeptides of the present invention include not only heterologous signal sequences, but also other heterologous functional regions. The fusion does not necessarily need to be direct, but may occur through linker sequences.

Moreover, fusion proteins may also be engineered to improve characteristics of the polypeptide of the present invention. For instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the

polypeptide to improve stability and persistence during purification from the host cell or subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide. The addition of peptide moieties to facilitate handling of polypeptides are familiar and routine techniques in the art.

Moreover, polypeptides of the present invention, including fragments, and specifically epitopes, can be combined with parts of the constant domain of immunoglobulins (IgG), resulting in chimeric polypeptides. These fusion proteins facilitate purification and show an increased half-life in vivo. One reported example describes chimeric proteins consisting of the first two domains of the human CD4-polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins. (EP A 394,827; Traunecker et al., Nature 331:84-86 (1988).) Fusion proteins having disulfide-linked dimeric structures (due to the IgG) can also be more efficient in binding and neutralizing other molecules, than the monomeric secreted protein or protein fragment alone. (Fountoulakis et al., J. Biochem. 270:3958-3964 (1995).)

Similarly, EP-A-O 464 533 (Canadian counterpart 2045869) discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, the Fc part in a fusion protein is beneficial in therapy and diagnosis, and thus can result in, for example, improved pharmacokinetic properties. (EP-A 0232 262.) Alternatively, deleting the Fc part after the fusion protein has been expressed, detected, and purified, would be desired. For example, the Fc portion may hinder therapy and diagnosis if the fusion protein is used as an antigen for immunizations. In drug discovery, for example, human proteins, such as hIL-5, have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists of hIL-5. (See, D. Bennett et al., J. Molecular Recognition 8:52-58 (1995); K. Johanson et al., J. Biol. Chem. 270:9459-9471 (1995).)

Moreover, the polypeptides of the present invention can be fused to marker sequences, such as a peptide which facilitates purification of the fused polypeptide. In preferred embodiments, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311), among others, many of which are commercially available. As described in Gentz et al., Proc. Natl. Acad. Sci. USA 86:821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein. Another peptide tag useful for purification, the "HA" tag, corresponds to an epitope derived from the influenza hemagglutinin protein. (Wilson et al., Cell 37:767 (1984).)

Thus, any of these above fusions can be engineered using the polynucleotides or the polypeptides of the present invention.

## Vectors, Host Cells, and Protein Production

The present invention also relates to vectors containing the polynucleotide of the present invention, host cells, and the production of polypeptides by recombinant techniques. The vector may be, for example, a phage, plasmid, viral, or retroviral vector. Retroviral vectors may be replication competent or replication defective. In the latter case, viral propagation generally will occur only in complementing host cells.

The polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector is introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is a virus, it may be packaged in vitro using an appropriate packaging cell line and then transduced into host cells.

The polynucleotide insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the E. coli lac, trp, phoA and tac promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled artisan. The expression constructs will further contain sites for transcription initiation, termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of the transcripts expressed by the constructs will preferably include a translation initiating codon at the beginning and a termination codon (UAA, UGA or UAG) appropriately positioned at the end of the polypeptide to be translated.

As indicated, the expression vectors will preferably include at least one selectable marker. Such markers include dihydrofolate reductase, G418 or neomycin resistance for eukaryotic cell culture and tetracycline, kanamycin or ampicillin resistance genes for culturing in E. coli and other bacteria. Representative examples of appropriate hosts include, but are not limited to, bacterial cells, such as E. coli, Streptomyces and Salmonella typhimurium cells; fungal cells, such as yeast cells; insect cells such as Drosophila S2 and Spodoptera Sf9 cells; animal cells such as CHO, COS, 293, and Bowes melanoma cells; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

Among vectors preferred for use in bacteria include pQE70, pQE60 and pQE-9, available from QIAGEN, Inc.; pBluescript vectors, Phagescript vectors, pNH8A, pNH16a, pNH18A, pNH46A, available from Stratagene Cloning Systems, Inc.; and ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia Biotech, Inc. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1

and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Other suitable vectors will be readily apparent to the skilled artisan.

Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, or other methods. Such methods are described in many standard laboratory manuals, such as Davis et al., Basic Methods In Molecular Biology (1986). It is specifically contemplated that the polypeptides of the present invention may in fact be expressed by a host cell lacking a recombinant vector.

A polypeptide of this invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification.

Polypeptides of the present invention, and preferably the secreted form, can also be recovered from: products purified from natural sources, including bodily fluids, tissues and cells, whether directly isolated or cultured; products of chemical synthetic procedures; and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect, and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes. Thus, it is well known in the art that the N-terminal methionine encoded by the translation initiation codon generally is removed with high efficiency from any protein after translation in all eukaryotic cells. While the N-terminal methionine on most proteins also is efficiently removed in most prokaryotes, for some proteins, this prokaryotic removal process is inefficient, depending on the nature of the amino acid to which the N-terminal methionine is covalently linked.

## Uses of the Polynucleotides

Each of the polynucleotides identified herein can be used in numerous ways as reagents. The following description should be considered exemplary and utilizes known techniques.

The polynucleotides of the present invention are useful for chromosome identification. There exists an ongoing need to identify new chromosome markers,

since few chromosome marking reagents, based on actual sequence data (repeat polymorphisms), are presently available. Each polynucleotide of the present invention can be used as a chromosome marker.

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the sequences shown in SEQ ID NO:X. Primers can be selected using computer analysis so that primers do not span more than one predicted exon in the genomic DNA. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the SEQ ID NO:X will yield an amplified fragment.

Similarly, somatic hybrids provide a rapid method of PCR mapping the polynucleotides to particular chromosomes. Three or more clones can be assigned per day using a single thermal cycler. Moreover, sublocalization of the polynucleotides can be achieved with panels of specific chromosome fragments. Other gene mapping strategies that can be used include in situ hybridization, prescreening with labeled flow-sorted chromosomes, and preselection by hybridization to construct chromosome specific-cDNA libraries.

Precise chromosomal location of the polynucleotides can also be achieved using fluorescence in situ hybridization (FISH) of a metaphase chromosomal spread. This technique uses polynucleotides as short as 500 or 600 bases; however, polynucleotides 2,000-4,000 bp are preferred. For a review of this technique, see Verma et al., "Human Chromosomes: a Manual of Basic Techniques," Pergamon Press, New York (1988).

For chromosome mapping, the polynucleotides can be used individually (to mark a single chromosome or a single site on that chromosome) or in panels (for marking multiple sites and/or multiple chromosomes). Preferred polynucleotides correspond to the noncoding regions of the cDNAs because the coding sequences are more likely conserved within gene families, thus increasing the chance of cross hybridization during chromosomal mapping.

Once a polynucleotide has been mapped to a precise chromosomal location, the physical position of the polynucleotide can be used in linkage analysis. Linkage analysis establishes coinheritance between a chromosomal location and presentation of a particular disease. (Disease mapping data are found, for example, in V. McKusick, Mendelian Inheritance in Man (available on line through Johns Hopkins University Welch Medical Library) .) Assuming 1 megabase mapping resolution and one gene per 20 kb, a cDNA precisely localized to a chromosomal region associated with the disease could be one of 50-500 potential causative genes.

Thus, once coinheritance is established, differences in the polynucleotide and the corresponding gene between affected and unaffected individuals can be examined. First, visible structural alterations in the chromosomes, such as deletions or translocations, are examined in chromosome spreads or by PCR. If no structural alterations exist, the presence of point mutations are ascertained. Mutations observed in some or all affected individuals, but not in normal individuals, indicates that the mutation may cause the disease. However, complete sequencing of the polypeptide and the corresponding gene from several normal individuals is required to distinguish the mutation from a polymorphism. If a new polymorphism is identified, this polymorphic polypeptide can be used for further linkage analysis.

Furthermore, increased or decreased expression of the gene in affected individuals as compared to unaffected individuals can be assessed using polynucleotides of the present invention. Any of these alterations (altered expression, chromosomal rearrangement, or mutation) can be used as a diagnostic or prognostic marker.

In addition to the foregoing, a polynucleotide can be used to control gene expression through triple helix formation or antisense DNA or RNA. Both methods rely on binding of the polynucleotide to DNA or RNA. For these techniques, preferred polynucleotides are usually 20 to 40 bases in length and complementary to either the region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 241:456 (1988); and Dervan et al., Science 251:1360 (1991) ) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxy-nucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988).) Triple helix formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques are effective in model systems, and the information disclosed herein can be used to design antisense or triple helix polynucleotides in an effort to treat disease.

Polynucleotides of the present invention are also useful in gene therapy. One goal of gene therapy is to insert a normal gene into an organism having a defective gene, in an effort to correct the genetic defect. The polynucleotides disclosed in the present invention offer a means of targeting such genetic defects in a highly accurate manner. Another goal is to insert a new gene that was not present in the host genome, thereby producing a new trait in the host cell.

The polynucleotides are also useful for identifying individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In

this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identifying personnel. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The polynucleotides of the present invention can be used as additional DNA markers for RFLP.

The polynucleotides of the present invention can also be used as an alternative to RFLP, by determining the actual base-by-base DNA sequence of selected portions of an individual's genome. These sequences can be used to prepare PCR primers for amplifying and isolating such selected DNA, which can then be sequenced. Using this technique, individuals can be identified because each individual will have a unique set of DNA sequences. Once an unique ID database is established for an individual, positive identification of that individual, living or dead, can be made from extremely small tissue samples.

Forensic biology also benefits from using DNA-based identification techniques as disclosed herein. DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, semen, etc., can be amplified using PCR. In one prior art technique, gene sequences amplified from polymorphic loci, such as DQa class II HLA gene, are used in forensic biology to identify individuals. (Erlich, H., PCR Technology, Freeman and Co. (1992).) Once these specific polymorphic loci are amplified, they are digested with one or more restriction enzymes, yielding an identifying set of bands on a Southern blot probed with DNA corresponding to the DQa class II HLA gene. Similarly, polynucleotides of the present invention can be used as polymorphic markers for forensic purposes.

There is also a need for reagents capable of identifying the source of a particular tissue. Such need arises, for example, in forensics when presented with tissue of unknown origin. Appropriate reagents can comprise, for example, DNA probes or primers specific to particular tissue prepared from the sequences of the present invention. Panels of such reagents can identify tissue by species and/or by organ type. In a similar fashion, these reagents can be used to screen tissue cultures for contamination.

In the very least, the polynucleotides of the present invention can be used as molecular weight markers on Southern gels, as diagnostic probes for the presence of a specific mRNA in a particular cell type, as a probe to "subtract-out" known sequences in the process of discovering novel polynucleotides, for selecting and making oligomers for attachment to a "gene chip" or other support, to raise anti-DNA antibodies using DNA immunization techniques, and as an antigen to elicit an immune response.

## Uses of the Polypeptides

Each of the polypeptides identified herein can be used in numerous ways. The following description should be considered exemplary and utilizes known techniques.

A polypeptide of the present invention can be used to assay protein levels in a biological sample using antibody-based techniques. For example, protein expression in tissues can be studied with classical immunohistological methods. (Jalkanen, M., et al., J. Cell. Biol. 101:976-985 (1985); Jalkanen, M., et al., J. Cell . Biol. 105:3087-3096 (1987).) Other antibody-based methods useful for detecting protein gene expression include immunoassays, such as the enzyme linked immunosorbent assay (ELISA) and the radioimmunoassay (RIA). Suitable antibody assay labels are known in the art and include enzyme labels, such as, glucose oxidase, and radioisotopes, such as iodine (125I, 121I), carbon (14C), sulfur (35S), tritium (3H), indium (112In), and technetium (99mTc), and fluorescent labels, such as fluorescein and rhodamine, and biotin.

In addition to assaying secreted protein levels in a biological sample, proteins can also be detected in vivo by imaging. Antibody labels or markers for in vivo imaging of protein include those detectable by X-radiography, NMR or ESR. For X-radiography, suitable labels include radioisotopes such as barium or cesium, which emit detectable radiation but are not overtly harmful to the subject. Suitable markers for NMR and ESR include those with a detectable characteristic spin, such as deuterium, which may be incorporated into the antibody by labeling of nutrients for the relevant hybridoma.

A protein-specific antibody or antibody fragment which has been labeled with an appropriate detectable imaging moiety, such as a radioisotope (for example, 131I, 112In, 99mTc), a radio-opaque substance, or a material detectable by nuclear magnetic resonance, is introduced (for example, parenterally, subcutaneously, or intraperitoneally) into the mammal. It will be understood in the art that the size of the subject and the imaging system used will determine the quantity of imaging moiety needed to produce diagnostic images. In the case of a radioisotope moiety, for a human subject, the quantity of radioactivity injected will normally range from about 5 to 20 millicuries of 99mTc. The labeled antibody or antibody fragment will then preferentially accumulate at the location of cells which contain the specific protein. In vivo tumor imaging is described in S.W. Burchiel et al., "Immunopharmacokinetics of Radiolabeled Antibodies and Their Fragments." (Chapter 13 in Tumor Imaging: The Radiochemical Detection of Cancer, S.W. Burchiel and B. A. Rhodes, eds., Masson Publishing Inc. (1982).)

Thus, the invention provides a diagnostic method of a disorder, which involves (a) assaying the expression of a polypeptide of the present invention in cells or body fluid of an individual; (b) comparing the level of gene expression with a standard gene expression level, whereby an increase or decrease in the assayed polypeptide gene expression level compared to the standard expression level is indicative of a disorder.

Moreover, polypeptides of the present invention can be used to treat disease. For example, patients can be administered a polypeptide of the present invention in an effort to replace absent or decreased levels of the polypeptide (e.g., insulin), to supplement absent or decreased levels of a different polypeptide (e.g., hemoglobin S for hemoglobin B), to inhibit the activity of a polypeptide (e.g., an oncogene), to activate the activity of a polypeptide (e.g., by binding to a receptor), to reduce the activity of a membrane bound receptor by competing with it for free ligand (e.g., soluble TNF receptors used in reducing inflammation), or to bring about a desired response (e.g., blood vessel growth).

Similarly, antibodies directed to a polypeptide of the present invention can also be used to treat disease. For example, administration of an antibody directed to a polypeptide of the present invention can bind and reduce overproduction of the polypeptide. Similarly, administration of an antibody can activate the polypeptide, such as by binding to a polypeptide bound to a membrane (receptor).

At the very least, the polypeptides of the present invention can be used as molecular weight markers on SDS-PAGE gels or on molecular sieve gel filtration columns using methods well known to those of skill in the art. Polypeptides can also be used to raise antibodies, which in turn are used to measure protein expression from a recombinant cell, as a way of assessing transformation of the host cell. Moreover, the polypeptides of the present invention can be used to test the following biological activities.

**Biological Activities**

The polynucleotides and polypeptides of the present invention can be used in assays to test for one or more biological activities. If these polynucleotides and polypeptides do exhibit activity in a particular assay, it is likely that these molecules may be involved in the diseases associated with the biological activity. Thus, the polynucleotides and polypeptides could be used to treat the associated disease.

**Immune Activity**

A polypeptide or polynucleotide of the present invention may be useful in treating deficiencies or disorders of the immune system, by activating or inhibiting the

proliferation, differentiation, or mobilization (chemotaxis) of immune cells. Immune cells develop through a process called hematopoiesis, producing myeloid (platelets, red blood cells, neutrophils, and macrophages) and lymphoid (B and T lymphocytes) cells from pluripotent stem cells. The etiology of these immune deficiencies or disorders may be genetic, somatic, such as cancer or some autoimmune disorders, acquired (e.g., by chemotherapy or toxins), or infectious. Moreover, a polynucleotide or polypeptide of the present invention can be used as a marker or detector of a particular immune system disease or disorder.

A polynucleotide or polypeptide of the present invention may be useful in treating or detecting deficiencies or disorders of hematopoietic cells. A polypeptide or polynucleotide of the present invention could be used to increase differentiation and proliferation of hematopoietic cells, including the pluripotent stem cells, in an effort to treat those disorders associated with a decrease in certain (or many) types hematopoietic cells. Examples of immunologic deficiency syndromes include, but are not limited to: blood protein disorders (e.g. agammaglobulinemia, dysgammaglobulinemia), ataxia telangiectasia, common variable immunodeficiency, Digeorge Syndrome, HIV infection, HTLV-BLV infection, leukocyte adhesion deficiency syndrome, lymphopenia, phagocyte bactericidal dysfunction, severe combined immunodeficiency (SCIDs), Wiskott-Aldrich Disorder, anemia, thrombocytopenia, or hemoglobinuria.

Moreover, a polypeptide or polynucleotide of the present invention could also be used to modulate hemostatic (the stopping of bleeding) or thrombolytic activity (clot formation). For example, by increasing hemostatic or thrombolytic activity, a polynucleotide or polypeptide of the present invention could be used to treat blood coagulation disorders (e.g., afibrinogenemia, factor deficiencies), blood platelet disorders (e.g. thrombocytopenia), or wounds resulting from trauma, surgery, or other causes. Alternatively, a polynucleotide or polypeptide of the present invention that can decrease hemostatic or thrombolytic activity could be used to inhibit or dissolve clotting. These molecules could be important in the treatment of heart attacks (infarction), strokes, or scarring.

A polynucleotide or polypeptide of the present invention may also be useful in treating or detecting autoimmune disorders. Many autoimmune disorders result from inappropriate recognition of self as foreign material by immune cells. This inappropriate recognition results in an immune response leading to the destruction of the host tissue. Therefore, the administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing autoimmune disorders.

Examples of autoimmune disorders that can be treated or detected by the present invention include, but are not limited to: Addison's Disease, hemolytic anemia, antiphospholipid syndrome, rheumatoid arthritis, dermatitis, allergic encephalomyelitis, glomerulonephritis, Goodpasture's Syndrome, Graves' Disease, Multiple Sclerosis, Myasthenia Gravis, Neuritis, Ophthalmia, Bullous Pemphigoid, Pemphigus, Polyendocrinopathies, Purpura, Reiter's Disease, Stiff-Man Syndrome, Autoimmune Thyroiditis, Systemic Lupus Erythematosus, Autoimmune Pulmonary Inflammation, Guillain-Barre Syndrome, insulin dependent diabetes mellitis, and autoimmune inflammatory eye disease.

Similarly, allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems, may also be treated by a polypeptide or polynucleotide of the present invention. Moreover, these molecules can be used to treat anaphylaxis, hypersensitivity to an antigenic molecule, or blood group incompatibility.

A polynucleotide or polypeptide of the present invention may also be used to treat and/or prevent organ rejection or graft-versus-host disease (GVHD). Organ rejection occurs by host immune cell destruction of the transplanted tissue through an immune response. Similarly, an immune response is also involved in GVHD, but, in this case, the foreign transplanted immune cells destroy the host tissues. The administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing organ rejection or GVHD.

Similarly, a polypeptide or polynucleotide of the present invention may also be used to modulate inflammation. For example, the polypeptide or polynucleotide may inhibit the proliferation and differentiation of cells involved in an inflammatory response. These molecules can be used to treat inflammatory conditions, both chronic and acute conditions, including inflammation associated with infection (e.g., septic shock, sepsis, or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine induced lung injury, inflammatory bowel disease, Crohn's disease, or resulting from over production of cytokines (e.g., TNF or IL-1.)

**Hyperproliferative Disorders**

A polypeptide or polynucleotide can be used to treat or detect hyperproliferative disorders, including neoplasms. A polypeptide or polynucleotide of the present invention may inhibit the proliferation of the disorder through direct or indirect

interactions. Alternatively, a polypeptide or polynucleotide of the present invention may proliferate other cells which can inhibit the hyperproliferative disorder.

For example, by increasing an immune response, particularly increasing antigenic qualities of the hyperproliferative disorder or by proliferating, differentiating, or mobilizing T-cells, hyperproliferative disorders can be treated. This immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, decreasing an immune response may also be a method of treating hyperproliferative disorders, such as a chemotherapeutic agent.

Examples of hyperproliferative disorders that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but are not limited to neoplasms located in the: abdomen, bone, breast, digestive system, liver, pancreas, peritoneum, endocrine glands (adrenal, parathyroid, pituitary, testicles, ovary, thymus, thyroid), eye, head and neck, nervous (central and peripheral), lymphatic system, pelvic, skin, soft tissue, spleen, thoracic, and urogenital.

Similarly, other hyperproliferative disorders can also be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of such hyperproliferative disorders include, but are not limited to: hypergammaglobulinemia, lymphoproliferative disorders, paraproteinemias, purpura, sarcoidosis, Sezary Syndrome, Waldenstron's Macroglobulinemia, Gaucher's Disease, histiocytosis, and any other hyperproliferative disease, besides neoplasia, located in an organ system listed above.

## Infectious Disease

A polypeptide or polynucleotide of the present invention can be used to treat or detect infectious agents. For example, by increasing the immune response, particularly increasing the proliferation and differentiation of B and/or T cells, infectious diseases may be treated. The immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, the polypeptide or polynucleotide of the present invention may also directly inhibit the infectious agent, without necessarily eliciting an immune response.

Viruses are one example of an infectious agent that can cause disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of viruses, include, but are not limited to the following DNA and RNA viral families: Arbovirus, Adenoviridae, Arenaviridae, Arterivirus, Birnaviridae, Bunyaviridae, Caliciviridae, Circoviridae, Coronaviridae, Flaviviridae, Hepadnaviridae (Hepatitis), Herpesviridae (such as, Cytomegalovirus, Herpes

Simplex, Herpes Zoster), Mononegavirus (e.g., Paramyxoviridae, Morbillivirus, Rhabdoviridae), Orthomyxoviridae (e.g., Influenza), Papovaviridae, Parvoviridae, Picornaviridae, Poxviridae (such as Smallpox or Vaccinia), Reoviridae (e.g., Rotavirus), Retroviridae (HTLV-I, HTLV-II, Lentivirus), and Togaviridae (e.g., Rubivirus). Viruses falling within these families can cause a variety of diseases or symptoms, including, but not limited to: arthritis, bronchiollitis, encephalitis, eye infections (e.g., conjunctivitis, keratitis), chronic fatigue syndrome, hepatitis (A, B, C, E, Chronic Active, Delta), meningitis, opportunistic infections (e.g., AIDS), pneumonia, Burkitt's Lymphoma, chickenpox , hemorrhagic fever, Measles, Mumps, Parainfluenza, Rabies, the common cold, Polio, leukemia, Rubella, sexually transmitted diseases, skin diseases (e.g., Kaposi's, warts), and viremia. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Similarly, bacterial or fungal agents that can cause disease or symptoms and that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following Gram-Negative and Gram-positive bacterial families and fungi: Actinomycetales (e.g., Corynebacterium, Mycobacterium, Norcardia), Aspergillosis, Bacillaceae (e.g., Anthrax, Clostridium), Bacteroidaceae, Blastomycosis, Bordetella, Borrelia, Brucellosis, Candidiasis, Campylobacter, Coccidioidomycosis, Cryptococcosis, Dermatocycoses, Enterobacteriaceae (Klebsiella, Salmonella, Serratia, Yersinia), Erysipelothrix, Helicobacter, Legionellosis, Leptospirosis, Listeria, Mycoplasmatales, Neisseriaceae (e.g., Acinetobacter, Gonorrhea, Menigococcal), Pasteurellacea Infections (e.g., Actinobacillus, Heamophilus, Pasteurella), Pseudomonas, Rickettsiaceae, Chlamydiaceae, Syphilis, and Staphylococcal. These bacterial or fungal families can cause the following diseases or symptoms, including, but not limited to: bacteremia, endocarditis, eye infections (conjunctivitis, tuberculosis, uveitis), gingivitis, opportunistic infections (e.g., AIDS related infections), paronychia, prosthesis-related infections, Reiter's Disease, respiratory tract infections, such as Whooping Cough or Empyema, sepsis, Lyme Disease, Cat-Scratch Disease, Dysentery, Paratyphoid Fever, food poisoning, Typhoid, pneumonia, Gonorrhea, meningitis, Chlamydia, Syphilis, Diphtheria, Leprosy, Paratuberculosis, Tuberculosis, Lupus, Botulism, gangrene, tetanus, impetigo, Rheumatic Fever, Scarlet Fever, sexually transmitted diseases, skin diseases (e.g., cellulitis, dermatocycoses), toxemia, urinary tract infections, wound infections. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Moreover, parasitic agents causing disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following families: Amebiasis, Babesiosis, Coccidiosis, Cryptosporidiosis, Dientamoebiasis, Dourine, Ectoparasitic, Giardiasis, Helminthiasis, Leishmaniasis, Theileriasis, Toxoplasmosis, Trypanosomiasis, and Trichomonas. These parasites can cause a variety of diseases or symptoms, including, but not limited to: Scabies, Trombiculiasis, eye infections, intestinal disease (e.g., dysentery, giardiasis), liver disease, lung disease, opportunistic infections (e.g., AIDS related), Malaria, pregnancy complications, and toxoplasmosis. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Preferably, treatment using a polypeptide or polynucleotide of the present invention could either be by administering an effective amount of a polypeptide to the patient, or by removing cells from the patient, supplying the cells with a polynucleotide of the present invention, and returning the engineered cells to the patient (ex vivo therapy). Moreover, the polypeptide or polynucleotide of the present invention can be used as an antigen in a vaccine to raise an immune response against infectious disease.

**Regeneration**

A polynucleotide or polypeptide of the present invention can be used to differentiate, proliferate, and attract cells, leading to the regeneration of tissues. (See, Science 276:59-87 (1997).) The regeneration of tissues could be used to repair, replace, or protect tissue damaged by congenital defects, trauma (wounds, burns, incisions, or ulcers), age, disease (e.g. osteoporosis, osteocarthritis, periodontal disease, liver failure), surgery, including cosmetic plastic surgery, fibrosis, reperfusion injury, or systemic cytokine damage.

Tissues that could be regenerated using the present invention include organs (e.g., pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac), vascular (including vascular endothelium), nervous, hematopoietic, and skeletal (bone, cartilage, tendon, and ligament) tissue. Preferably, regeneration occurs without or decreased scarring. Regeneration also may include angiogenesis.

Moreover, a polynucleotide or polypeptide of the present invention may increase regeneration of tissues difficult to heal. For example, increased tendon/ligament regeneration would quicken recovery time after damage. A polynucleotide or polypeptide of the present invention could also be used prophylactically in an effort to avoid damage. Specific diseases that could be treated include of tendinitis, carpal tunnel syndrome, and other tendon or ligament defects. A further example of tissue

regeneration of non-healing wounds includes pressure ulcers, ulcers associated with vascular insufficiency, surgical, and traumatic wounds.

Similarly, nerve and brain tissue could also be regenerated by using a polynucleotide or polypeptide of the present invention to proliferate and differentiate nerve cells. Diseases that could be treated using this method include central and peripheral nervous system diseases, neuropathies, or mechanical and traumatic disorders (e.g., spinal cord disorders, head trauma, cerebrovascular disease, and stoke). Specifically, diseases associated with peripheral nerve injuries, peripheral neuropathy (e.g., resulting from chemotherapy or other medical therapies), localized neuropathies, and central nervous system diseases (e.g., Alzheimer's disease, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome), could all be treated using the polynucleotide or polypeptide of the present invention.

## Chemotaxis

A polynucleotide or polypeptide of the present invention may have chemotaxis activity. A chemotaxic molecule attracts or mobilizes cells (e.g., monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells) to a particular site in the body, such as inflammation, infection, or site of hyperproliferation. The mobilized cells can then fight off and/or heal the particular trauma or abnormality.

A polynucleotide or polypeptide of the present invention may increase chemotaxic activity of particular cells. These chemotactic molecules can then be used to treat inflammation, infection, hyperproliferative disorders, or any immune system disorder by increasing the number of cells targeted to a particular location in the body. For example, chemotaxic molecules can be used to treat wounds and other trauma to tissues by attracting immune cells to the injured location. Chemotactic molecules of the present invention can also attract fibroblasts, which can be used to treat wounds.

It is also contemplated that a polynucleotide or polypeptide of the present invention may inhibit chemotactic activity. These molecules could also be used to treat disorders. Thus, a polynucleotide or polypeptide of the present invention could be used as an inhibitor of chemotaxis.

## Binding Activity

A polypeptide of the present invention may be used to screen for molecules that bind to the polypeptide or for molecules to which the polypeptide binds. The binding of the polypeptide and the molecule may activate (agonist), increase, inhibit

(antagonist), or decrease activity of the polypeptide or the molecule bound. Examples of such molecules include antibodies, oligonucleotides, proteins (e.g., receptors),or small molecules.

Preferably, the molecule is closely related to the natural ligand of the polypeptide, e.g., a fragment of the ligand, or a natural substrate, a ligand, a structural or functional mimetic. (See, Coligan et al., Current Protocols in Immunology 1(2):Chapter 5 (1991).) Similarly, the molecule can be closely related to the natural receptor to which the polypeptide binds, or at least, a fragment of the receptor capable of being bound by the polypeptide (e.g., active site). In either case, the molecule can be rationally designed using known techniques.

Preferably, the screening for these molecules involves producing appropriate cells which express the polypeptide, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or *E. coli*. Cells expressing the polypeptide (or cell membrane containing the expressed polypeptide) are then preferably contacted with a test compound potentially containing the molecule to observe binding, stimulation, or inhibition of activity of either the polypeptide or the molecule.

The assay may simply test binding of a candidate compound to the polypeptide, wherein binding is detected by a label, or in an assay involving competition with a labeled competitor. Further, the assay may test whether the candidate compound results in a signal generated by binding to the polypeptide.

Alternatively, the assay can be carried out using cell-free preparations, polypeptide/molecule affixed to a solid support, chemical libraries, or natural product mixtures. The assay may also simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide, measuring polypeptide/molecule activity or binding, and comparing the polypeptide/molecule activity or binding to a standard.

Preferably, an ELISA assay can measure polypeptide level or activity in a sample (e.g., biological sample) using a monoclonal or polyclonal antibody. The antibody can measure polypeptide level or activity by either binding, directly or indirectly, to the polypeptide or by competing with the polypeptide for a substrate.

All of these above assays can be used as diagnostic or prognostic markers. The molecules discovered using these assays can be used to treat disease or to bring about a particular result in a patient (e.g., blood vessel growth) by activating or inhibiting the polypeptide/molecule. Moreover, the assays can discover agents which may inhibit or enhance the production of the polypeptide from suitably manipulated cells or tissues.

Therefore, the invention includes a method of identifying compounds which bind to a polypeptide of the invention comprising the steps of: (a) incubating a candidate binding compound with a polypeptide of the invention; and (b) determining if binding has occurred. Moreover, the invention includes a method of identifying agonists/antagonists comprising the steps of: (a) incubating a candidate compound with a polypeptide of the invention, (b) assaying a biological activity , and (b) determining if a biological activity of the polypeptide has been altered.

## Other Activities

A polypeptide or polynucleotide of the present invention may also increase or decrease the differentiation or proliferation of embryonic stem cells, besides, as discussed above, hematopoietic lineage.

A polypeptide or polynucleotide of the present invention may also be used to modulate mammalian characteristics, such as body height, weight, hair color, eye color, skin, percentage of adipose tissue, pigmentation, size, and shape (e.g., cosmetic surgery). Similarly, a polypeptide or polynucleotide of the present invention may be used to modulate mammalian metabolism affecting catabolism, anabolism, processing, utilization, and storage of energy.

A polypeptide or polynucleotide of the present invention may be used to change a mammal's mental state or physical state by influencing biorhythms, caricadic rhythms, depression (including depressive disorders), tendency for violence, tolerance for pain, reproductive capabilities (preferably by Activin or Inhibin-like activity), hormonal or endocrine levels, appetite, libido, memory, stress, or other cognitive qualities.

A polypeptide or polynucleotide of the present invention may also be used as a food additive or preservative, such as to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional components.

## Other Preferred Embodiments

Other preferred embodiments of the claimed invention include an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 50 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of

positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Clone Sequence and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Start Codon and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Similarly preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 150 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

Further preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 500 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

A further preferred embodiment is a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the nucleotide sequence of SEQ ID NO:X beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence of SEQ ID NO:X.

Also preferred is an isolated nucleic acid molecule which hybridizes under stringent hybridization conditions to a nucleic acid molecule, wherein said nucleic acid molecule which hybridizes does not hybridize under stringent hybridization conditions to a nucleic acid molecule having a nucleotide sequence consisting of only A residues or of only T residues.

Also preferred is a composition of matter comprising a DNA molecule which comprises a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the material deposited with the American Type

Culture Collection and given the ATCC Deposit Number shown in Table 1 for said cDNA Clone Identifier.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in the nucleotide sequence of a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the deposit given the ATCC Deposit Number shown in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said sequence of at least 50 contiguous nucleotides is included in the nucleotide sequence of the complete open reading frame sequence encoded by said human cDNA clone.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 150 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 500 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is a method for detecting in a biological sample a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method comprises a step of comparing a nucleotide sequence of at least one nucleic acid molecule in said sample with a sequence selected from said group and determining whether the sequence of said nucleic acid molecule in said sample is at least 95% identical to said selected sequence.

Also preferred is the above method wherein said step of comparing sequences comprises determining the extent of nucleic acid hybridization between nucleic acid molecules in said sample and a nucleic acid molecule comprising said sequence selected from said group. Similarly, also preferred is the above method wherein said step of comparing sequences is performed by comparing the nucleotide sequence determined from a nucleic acid molecule in said sample with said sequence selected from said group. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

A further preferred embodiment is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting nucleic acid molecules in said sample, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

The method for identifying the species, tissue or cell type of a biological sample can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject nucleic acid molecules, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

The method for diagnosing a pathological condition can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a composition of matter comprising isolated nucleic acid molecules wherein the nucleotide sequences of said nucleic acid molecules comprise a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1.

Also preferred is a polypeptide, wherein said sequence of contiguous amino acids is included in the amino acid sequence of SEQ ID NO:Y in the range of positions beginning with the residue at about the position of the First Amino Acid of the Secreted Portion and ending with the residue at about the Last Amino Acid of the Open Reading Frame as set forth for SEQ ID NO:Y in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the complete amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is a polypeptide wherein said sequence of contiguous amino acids is included in the amino acid sequence of a secreted portion of the secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is an isolated antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method for detecting in a biological sample a polypeptide comprising an amino acid sequence which is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method comprises a step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group and determining whether the sequence of said polypeptide molecule in said sample is at least 90% identical to said sequence of at least 10 contiguous amino acids.

Also preferred is the above method wherein said step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group comprises determining the extent of specific binding of polypeptides in said sample to an antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method wherein said step of comparing sequences is performed by comparing the amino acid sequence determined from a polypeptide molecule in said sample with said sequence selected from said group.

Also preferred is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting polypeptide molecules in said sample, if any, comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method for identifying the species, tissue or cell type of a biological sample, which method comprises a step of detecting polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the above group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

In any of these methods, the step of detecting said polypeptide molecules includes using an antibody.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a nucleotide sequence encoding a polypeptide wherein said polypeptide comprises an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said nucleotide sequence encoding a polypeptide has been optimized for expression of said polypeptide in a prokaryotic host.

Also preferred is an isolated nucleic acid molecule, wherein said polypeptide comprises an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method of making a recombinant vector comprising inserting any of the above isolated nucleic acid molecule into a vector. Also preferred is the recombinant vector produced by this method. Also preferred is a method of making a recombinant host cell comprising introducing the vector into a host cell, as well as the recombinant host cell produced by this method.

Also preferred is a method of making an isolated polypeptide comprising culturing this recombinant host cell under conditions such that said polypeptide is expressed and recovering said polypeptide. Also preferred is this method of making an isolated polypeptide, wherein said recombinant host cell is a eukaryotic cell and said polypeptide is a secreted portion of a human secreted protein comprising an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y beginning with the residue at the position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y wherein Y is an integer set forth in Table 1 and said position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y is defined in Table 1; and an amino acid sequence of a secreted portion of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The isolated polypeptide produced by this method is also preferred.

Also preferred is a method of treatment of an individual in need of an increased level of a secreted protein activity, which method comprises administering to such an individual a pharmaceutical composition comprising an amount of an isolated polypeptide, polynucleotide, or antibody of the claimed invention effective to increase the level of said protein activity in said individual.

Having generally described the invention, the same will be more readily understood by reference to the following examples, which are provided by way of illustration and are not intended as limiting.

## Examples

### Example 1: Isolation of a Selected cDNA Clone From the Deposited Sample

Each cDNA clone in a cited ATCC deposit is contained in a plasmid vector. Table 1 identifies the vectors used to construct the cDNA library from which each clone was isolated. In many cases, the vector used to construct the library is a phage vector from which a plasmid has been excised. The table immediately below correlates the related plasmid for each phage vector used in constructing the cDNA library. For example, where a particular clone is identified in Table 1 as being isolated in the vector "Lambda Zap," the corresponding deposited clone is in "pBluescript."

| Vector Used to Construct Library | Corresponding Deposited Plasmid |
|---|---|
| Lambda Zap | pBluescript (pBS) |
| Uni-Zap XR | pBluescript (pBS) |
| Zap Express | pBK |
| lafmid BA | plafmid BA |
| pSport1 | pSport1 |
| pCMVSport 2.0 | pCMVSport 2.0 |
| pCMVSport 3.0 | pCMVSport 3.0 |
| pCR®2.1 | pCR®2.1 |

Vectors Lambda Zap (U.S. Patent Nos. 5,128,256 and 5,286,636), Uni-Zap XR (U.S. Patent Nos. 5,128, 256 and 5,286,636), Zap Express (U.S. Patent Nos. 5,128,256 and 5,286,636), pBluescript (pBS) (Short, J. M. et al., Nucleic Acids Res. 16:7583-7600 (1988); Alting-Mees, M. A. and Short, J. M., Nucleic Acids Res. 17:9494 (1989)) and pBK (Alting-Mees, M. A. et al., Strategies 5:58-61 (1992)) are commercially available from Stratagene Cloning Systems, Inc., 11011 N. Torrey Pines Road, La Jolla, CA, 92037. pBS contains an ampicillin resistance gene and pBK contains a neomycin resistance gene. Both can be transformed into E. coli strain XL-1 Blue, also available from Stratagene. pBS comes in 4 forms SK+, SK-, KS+ and KS. The S and K refers to the orientation of the polylinker to the T7 and T3 primer sequences which flank the polylinker region ("S" is for SacI and "K" is for KpnI which are the first sites on each respective end of the linker). "+" or "-" refer to the orientation of the f1 origin of replication ("ori"), such that in one orientation, single stranded rescue initiated from the f1 ori generates sense strand DNA and in the other, antisense.

Vectors pSport1, pCMVSport 2.0 and pCMVSport 3.0, were obtained from Life Technologies, Inc., P. O. Box 6009, Gaithersburg, MD 20897. All Sport vectors contain an ampicillin resistance gene and may be transformed into E. coli strain

DH10B, also available from Life Technologies. (See, for instance, Gruber, C. E., et al., Focus 15:59 (1993).) Vector lafmid BA (Bento Soares, Columbia University, NY) contains an ampicillin resistance gene and can be transformed into E. coli strain XL-1 Blue. Vector pCR®2.1, which is available from Invitrogen, 1600 Faraday Avenue, Carlsbad, CA 92008, contains an ampicillin resistance gene and may be transformed into E. coli strain DH10B, available from Life Technologies. (See, for instance, Clark, J. M., Nuc. Acids Res. 16:9677-9686 (1988) and Mead, D. et al., Bio/Technology 9: (1991).) Preferably, a polynucleotide of the present invention does not comprise the phage vector sequences identified for the particular clone in Table 1, as well as the corresponding plasmid vector sequences designated above.

The deposited material in the sample assigned the ATCC Deposit Number cited in Table 1 for any given cDNA clone also may contain one or more additional plasmids, each comprising a cDNA clone different from that given clone. Thus, deposits sharing the same ATCC Deposit Number contain at least a plasmid for each cDNA clone identified in Table 1. Typically, each ATCC deposit sample cited in Table 1 comprises a mixture of approximately equal amounts (by weight) of about 50 plasmid DNAs, each containing a different cDNA clone; but such a deposit sample may include plasmids for more or less than 50 cDNA clones, up to about 500 cDNA clones.

Two approaches can be used to isolate a particular clone from the deposited sample of plasmid DNAs cited for that clone in Table 1. First, a plasmid is directly isolated by screening the clones using a polynucleotide probe corresponding to SEQ ID NO:X.

Particularly, a specific polynucleotide with 30-40 nucleotides is synthesized using an Applied Biosystems DNA synthesizer according to the sequence reported. The oligonucleotide is labeled, for instance, with $^{32}$P-$\gamma$-ATP using T4 polynucleotide kinase and purified according to routine methods. (E.g., Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring, NY (1982).) The plasmid mixture is transformed into a suitable host, as indicated above (such as XL-1 Blue (Stratagene)) using techniques known to those of skill in the art, such as those provided by the vector supplier or in related publications or patents cited above. The transformants are plated on 1.5% agar plates (containing the appropriate selection agent, e.g., ampicillin) to a density of about 150 transformants (colonies) per plate. These plates are screened using Nylon membranes according to routine methods for bacterial colony screening (e.g., Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Edit., (1989), Cold Spring Harbor Laboratory Press, pages 1.93 to 1.104), or other techniques known to those of skill in the art.

Alternatively, two primers of 17-20 nucleotides derived from both ends of the SEQ ID NO:X (i.e., within the region of SEQ ID NO:X bounded by the 5' NT and the 3' NT of the clone defined in Table 1) are synthesized and used to amplify the desired cDNA using the deposited cDNA plasmid as a template. The polymerase chain reaction is carried out under routine conditions, for instance, in 25 $\mu$l of reaction mixture with 0.5 ug of the above cDNA template. A convenient reaction mixture is 1.5-5 mM $MgCl_2$, 0.01% (w/v) gelatin, 20 $\mu$M each of dATP, dCTP, dGTP, dTTP, 25 pmol of each primer and 0.25 Unit of Taq polymerase. Thirty five cycles of PCR (denaturation at 94°C for 1 min; annealing at 55°C for 1 min; elongation at 72°C for 1 min) are performed with a Perkin-Elmer Cetus automated thermal cycler. The amplified product is analyzed by agarose gel electrophoresis and the DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the selected sequence by subcloning and sequencing the DNA product.

Several methods are available for the identification of the 5' or 3' non-coding portions of a gene which may not be present in the deposited clone. These methods include but are not limited to, filter probing, clone enrichment using specific probes, and protocols similar or identical to 5' and 3' "RACE" protocols which are well known in the art. For instance, a method similar to 5' RACE is available for generating the missing 5' end of a desired full-length transcript. (Fromont-Racine et al., Nucleic Acids Res. 21(7):1683-1684 (1993).)

Briefly, a specific RNA oligonucleotide is ligated to the 5' ends of a population of RNA presumably containing full-length gene RNA transcripts. A primer set containing a primer specific to the ligated RNA oligonucleotide and a primer specific to a known sequence of the gene of interest is used to PCR amplify the 5' portion of the desired full-length gene. This amplified product may then be sequenced and used to generate the full length gene.

This above method starts with total RNA isolated from the desired source, although poly-A+ RNA can be used. The RNA preparation can then be treated with phosphatase if necessary to eliminate 5' phosphate groups on degraded or damaged RNA which may interfere with the later RNA ligase step. The phosphatase should then be inactivated and the RNA treated with tobacco acid pyrophosphatase in order to remove the cap structure present at the 5' ends of messenger RNAs. This reaction leaves a 5' phosphate group at the 5' end of the cap cleaved RNA which can then be ligated to an RNA oligonucleotide using T4 RNA ligase.

This modified RNA preparation is used as a template for first strand cDNA synthesis using a gene specific oligonucleotide. The first strand synthesis reaction is

used as a template for PCR amplification of the desired 5' end using a primer specific to the ligated RNA oligonucleotide and a primer specific to the known sequence of the gene of interest. The resultant product is then sequenced and analyzed to confirm that the 5' end sequence belongs to the desired gene.

**Example 2: Isolation of Genomic Clones Corresponding to a Polynucleotide**

A human genomic P1 library (Genomic Systems, Inc.) is screened by PCR using primers selected for the cDNA sequence corresponding to SEQ ID NO:X., according to the method described in Example 1. (See also, Sambrook.)

**Example 3: Tissue Distribution of Polypeptide**

Tissue distribution of mRNA expression of polynucleotides of the present invention is determined using protocols for Northern blot analysis, described by, among others, Sambrook et al. For example, a cDNA probe produced by the method described in Example 1 is labeled with $P^{32}$ using the rediprime™ DNA labeling system (Amersham Life Science), according to manufacturer's instructions. After labeling, the probe is purified using CHROMA SPIN-100™ column (Clontech Laboratories, Inc.), according to manufacturer's protocol number PT1200-1. The purified labeled probe is then used to examine various human tissues for mRNA expression.

Multiple Tissue Northern (MTN) blots containing various human tissues (H) or human immune system tissues (IM) (Clontech) are examined with the labeled probe using ExpressHyb™ hybridization solution (Clontech) according to manufacturer's protocol number PT1190-1. Following hybridization and washing, the blots are mounted and exposed to film at -70°C overnight, and the films developed according to standard procedures.

**Example 4: Chromosomal Mapping of the Polynucleotides**

An oligonucleotide primer set is designed according to the sequence at the 5' end of SEQ ID NO:X. This primer preferably spans about 100 nucleotides. This primer set is then used in a polymerase chain reaction under the following set of conditions : 30 seconds, 95°C; 1 minute, 56°C; 1 minute, 70°C. This cycle is repeated 32 times followed by one 5 minute cycle at 70°C. Human, mouse, and hamster DNA is used as template in addition to a somatic cell hybrid panel containing individual chromosomes or chromosome fragments (Bios, Inc). The reactions is analyzed on

either 8% polyacrylamide gels or 3.5 % agarose gels. Chromosome mapping is determined by the presence of an approximately 100 bp PCR fragment in the particular somatic cell hybrid.

## Example 5: Bacterial Expression of a Polypeptide

A polynucleotide encoding a polypeptide of the present invention is amplified using PCR oligonucleotide primers corresponding to the 5' and 3' ends of the DNA sequence, as outlined in Example 1, to synthesize insertion fragments. The primers used to amplify the cDNA insert should preferably contain restriction sites, such as BamHI and XbaI, at the 5' end of the primers in order to clone the amplified product into the expression vector. For example, BamHI and XbaI correspond to the restriction enzyme sites on the bacterial expression vector pQE-9. (Qiagen, Inc., Chatsworth, CA). This plasmid vector encodes antibiotic resistance ($Amp^r$), a bacterial origin of replication (ori), an IPTG-regulatable promoter/operator (P/O), a ribosome binding site (RBS), a 6-histidine tag (6-His), and restriction enzyme cloning sites.

The pQE-9 vector is digested with BamHI and XbaI and the amplified fragment is ligated into the pQE-9 vector maintaining the reading frame initiated at the bacterial RBS. The ligation mixture is then used to transform the E. coli strain M15/rep4 (Qiagen, Inc.) which contains multiple copies of the plasmid pREP4, which expresses the lacI repressor and also confers kanamycin resistance ($Kan^r$). Transformants are identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies are selected. Plasmid DNA is isolated and confirmed by restriction analysis.

Clones containing the desired constructs are grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 ug/ml) and Kan (25 ug/ml). The O/N culture is used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells are grown to an optical density 600 ($O.D.^{600}$) of between 0.4 and 0.6. IPTG (Isopropyl-B-D-thiogalacto pyranoside) is then added to a final concentration of 1 mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene expression.

Cells are grown for an extra 3 to 4 hours. Cells are then harvested by centrifugation (20 mins at 6000Xg). The cell pellet is solubilized in the chaotropic agent 6 Molar Guanidine HCl by stirring for 3-4 hours at 4°C. The cell debris is removed by centrifugation, and the supernatant containing the polypeptide is loaded onto a nickel-nitrilo-tri-acetic acid ("Ni-NTA") affinity resin column (available from QIAGEN, Inc., *supra*). Proteins with a 6 x His tag bind to the Ni-NTA resin with high

affinity and can be purified in a simple one-step procedure (for details see: The QIAexpressionist (1995) QIAGEN, Inc., *supra*).

Briefly, the supernatant is loaded onto the column in 6 M guanidine-HCl, pH 8, the column is first washed with 10 volumes of 6 M guanidine-HCl, pH 8, then washed with 10 volumes of 6 M guanidine-HCl pH 6, and finally the polypeptide is eluted with 6 M guanidine-HCl, pH 5.

The purified protein is then renatured by dialyzing it against phosphate-buffered saline (PBS) or 50 mM Na-acetate, pH 6 buffer plus 200 mM NaCl. Alternatively, the protein can be successfully refolded while immobilized on the Ni-NTA column. The recommended conditions are as follows: renature using a linear 6M-1M urea gradient in 500 mM NaCl, 20% glycerol, 20 mM Tris/HCl pH 7.4, containing protease inhibitors. The renaturation should be performed over a period of 1.5 hours or more. After renaturation the proteins are eluted by the addition of 250 mM immidazole. Immidazole is removed by a final dialyzing step against PBS or 50 mM sodium acetate pH 6 buffer plus 200 mM NaCl. The purified protein is stored at 4° C or frozen at -80° C.

In addition to the above expression vector, the present invention further includes an expression vector comprising phage operator and promoter elements operatively linked to a polynucleotide of the present invention, called pHE4a. (ATCC Accession Number 209645, deposited on February 25, 1998.) This vector contains: 1) a neomycinphosphotransferase gene as a selection marker, 2) an E. coli origin of replication, 3) a T5 phage promoter sequence, 4) two lac operator sequences, 5) a Shine-Delgarno sequence, and 6) the lactose operon repressor gene (lacIq). The origin of replication (oriC) is derived from pUC19 (LTI, Gaithersburg, MD). The promoter sequence and operator sequences are made synthetically.

DNA can be inserted into the pHEa by restricting the vector with NdeI and XbaI, BamHI, XhoI, or Asp718, running the restricted product on a gel, and isolating the larger fragment (the stuffer fragment should be about 310 base pairs). The DNA insert is generated according to the PCR protocol described in Example 1, using PCR primers having restriction sites for NdeI (5' primer) and XbaI, BamHI, XhoI, or Asp718 (3' primer). The PCR insert is gel purified and restricted with compatible enzymes. The insert and vector are ligated according to standard protocols.

The engineered vector could easily be substituted in the above protocol to express protein in a bacterial system.

**Example 6: Purification of a Polypeptide from an Inclusion Body**

The following alternative method can be used to purify a polypeptide expressed in *E coli* when it is present in the form of inclusion bodies. Unless otherwise specified, all of the following steps are conducted at 4-10°C.

Upon completion of the production phase of the *E. coli* fermentation, the cell culture is cooled to 4-10°C and the cells harvested by continuous centrifugation at 15,000 rpm (Heraeus Sepatech). On the basis of the expected yield of protein per unit weight of cell paste and the amount of purified protein required, an appropriate amount of cell paste, by weight, is suspended in a buffer solution containing 100 mM Tris, 50 mM EDTA, pH 7.4. The cells are dispersed to a homogeneous suspension using a high shear mixer.

The cells are then lysed by passing the solution through a microfluidizer (Microfuidics, Corp. or APV Gaulin, Inc.) twice at 4000-6000 psi. The homogenate is then mixed with NaCl solution to a final concentration of 0.5 M NaCl, followed by centrifugation at 7000 xg for 15 min. The resultant pellet is washed again using 0.5M NaCl, 100 mM Tris, 50 mM EDTA, pH 7.4.

The resulting washed inclusion bodies are solubilized with 1.5 M guanidine hydrochloride (GuHCl) for 2-4 hours. After 7000 xg centrifugation for 15 min., the pellet is discarded and the polypeptide containing supernatant is incubated at 4°C overnight to allow further GuHCl extraction.

Following high speed centrifugation (30,000 xg) to remove insoluble particles, the GuHCl solubilized protein is refolded by quickly mixing the GuHCl extract with 20 volumes of buffer containing 50 mM sodium, pH 4.5, 150 mM NaCl, 2 mM EDTA by vigorous stirring. The refolded diluted protein solution is kept at 4°C without mixing for 12 hours prior to further purification steps.

To clarify the refolded polypeptide solution, a previously prepared tangential filtration unit equipped with 0.16 µm membrane filter with appropriate surface area (e.g., Filtron), equilibrated with 40 mM sodium acetate, pH 6.0 is employed. The filtered sample is loaded onto a cation exchange resin (e.g., Poros HS-50, Perseptive Biosystems). The column is washed with 40 mM sodium acetate, pH 6.0 and eluted with 250 mM, 500 mM, 1000 mM, and 1500 mM NaCl in the same buffer, in a stepwise manner. The absorbance at 280 nm of the effluent is continuously monitored. Fractions are collected and further analyzed by SDS-PAGE.

Fractions containing the polypeptide are then pooled and mixed with 4 volumes of water. The diluted sample is then loaded onto a previously prepared set of tandem columns of strong anion (Poros HQ-50, Perseptive Biosystems) and weak anion (Poros CM-20, Perseptive Biosystems) exchange resins. The columns are equilibrated with 40 mM sodium acetate, pH 6.0. Both columns are washed with 40 mM sodium acetate, pH 6.0, 200 mM NaCl. The CM-20 column is then eluted using a 10 column volume linear gradient ranging from 0.2 M NaCl, 50 mM sodium acetate, pH 6.0 to 1.0 M NaCl, 50 mM sodium acetate, pH 6.5. Fractions are collected under constant $A_{280}$ monitoring of the effluent. Fractions containing the polypeptide (determined, for instance, by 16% SDS-PAGE) are then pooled.

The resultant polypeptide should exhibit greater than 95% purity after the above refolding and purification steps. No major contaminant bands should be observed from Commassie blue stained 16% SDS-PAGE gel when 5 μg of purified protein is loaded. The purified protein can also be tested for endotoxin/LPS contamination, and typically the LPS content is less than 0.1 ng/ml according to LAL assays.

## Example 7: Cloning and Expression of a Polypeptide in a Baculovirus Expression System

In this example, the plasmid shuttle vector pA2 is used to insert a polynucleotide into a baculovirus to express a polypeptide. This expression vector contains the strong polyhedrin promoter of the *Autographa californica* nuclear polyhedrosis virus (AcMNPV) followed by convenient restriction sites such as BamHI, Xba I and Asp718. The polyadenylation site of the simian virus 40 ("SV40") is used for efficient polyadenylation. For easy selection of recombinant virus, the plasmid contains the beta-galactosidase gene from *E. coli* under control of a weak Drosophila promoter in the same orientation, followed by the polyadenylation signal of the polyhedrin gene. The inserted genes are flanked on both sides by viral sequences for cell-mediated homologous recombination with wild-type viral DNA to generate a viable virus that express the cloned polynucleotide.

Many other baculovirus vectors can be used in place of the vector above, such as pAc373, pVL941, and pAcIM1, as one skilled in the art would readily appreciate, as long as the construct provides appropriately located signals for transcription, translation, secretion and the like, including a signal peptide and an in-frame AUG as required. Such vectors are described, for instance, in Luckow et al., Virology 170:31-39 (1989).

Specifically, the cDNA sequence contained in the deposited clone, including the AUG initiation codon and the naturally associated leader sequence identified in Table 1, is amplified using the PCR protocol described in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the pA2 vector does not need a second signal peptide. Alternatively, the vector can be modified (pA2 GP) to include a baculovirus leader sequence, using the standard methods described in Summers et al., "A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures," Texas Agricultural Experimental Station Bulletin No. 1555 (1987).

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("Geneclean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

The plasmid is digested with the corresponding restriction enzymes and optionally, can be dephosphorylated using calf intestinal phosphatase, using routine procedures known in the art. The DNA is then isolated from a 1% agarose gel using a commercially available kit ("Geneclean" BIO 101 Inc., La Jolla, Ca.).

The fragment and the dephosphorylated plasmid are ligated together with T4 DNA ligase. *E. coli* HB101 or other suitable *E. coli* hosts such as XL-1 Blue (Stratagene Cloning Systems, La Jolla, CA) cells are transformed with the ligation mixture and spread on culture plates. Bacteria containing the plasmid are identified by digesting DNA from individual colonies and analyzing the digestion product by gel electrophoresis. The sequence of the cloned fragment is confirmed by DNA sequencing.

Five µg of a plasmid containing the polynucleotide is co-transfected with 1.0 µg of a commercially available linearized baculovirus DNA ("BaculoGold™ baculovirus DNA", Pharmingen, San Diego, CA), using the lipofection method described by Felgner et al., Proc. Natl. Acad. Sci. USA 84:7413-7417 (1987). One µg of BaculoGold™ virus DNA and 5 µg of the plasmid are mixed in a sterile well of a microtiter plate containing 50 µl of serum-free Grace's medium (Life Technologies Inc., Gaithersburg, MD). Afterwards, 10 µl Lipofectin plus 90 µl Grace's medium are added, mixed and incubated for 15 minutes at room temperature. Then the transfection mixture is added drop-wise to Sf9 insect cells (ATCC CRL 1711) seeded in a 35 mm tissue culture plate with 1 ml Grace's medium without serum. The plate is then incubated for 5 hours at 27° C. The transfection solution is then removed from the plate and 1 ml of Grace's insect medium supplemented with 10% fetal calf serum is added. Cultivation is then continued at 27° C for four days.

After four days the supernatant is collected and a plaque assay is performed, as described by Summers and Smith, *supra*. An agarose gel with "Blue Gal" (Life

Technologies Inc., Gaithersburg) is used to allow easy identification and isolation of gal-expressing clones, which produce blue-stained plaques. (A detailed description of a "plaque assay" of this type can also be found in the user's guide for insect cell culture and baculovirology distributed by Life Technologies Inc., Gaithersburg, page 9-10.) After appropriate incubation, blue stained plaques are picked with the tip of a micropipettor (e.g., Eppendorf). The agar containing the recombinant viruses is then resuspended in a microcentrifuge tube containing 200 µl of Grace's medium and the suspension containing the recombinant baculovirus is used to infect Sf9 cells seeded in 35 mm dishes. Four days later the supernatants of these culture dishes are harvested and then they are stored at 4° C.

To verify the expression of the polypeptide, Sf9 cells are grown in Grace's medium supplemented with 10% heat-inactivated FBS. The cells are infected with the recombinant baculovirus containing the polynucleotide at a multiplicity of infection ("MOI") of about 2. If radiolabeled proteins are desired, 6 hours later the medium is removed and is replaced with SF900 II medium minus methionine and cysteine (available from Life Technologies Inc., Rockville, MD). After 42 hours, 5 µCi of $^{35}$S-methionine and 5 µCi $^{35}$S-cysteine (available from Amersham) are added. The cells are further incubated for 16 hours and then are harvested by centrifugation. The proteins in the supernatant as well as the intracellular proteins are analyzed by SDS-PAGE followed by autoradiography (if radiolabeled).

Microsequencing of the amino acid sequence of the amino terminus of purified protein may be used to determine the amino terminal sequence of the produced protein.

**Example 8: Expression of a Polypeptide in Mammalian Cells**

The polypeptide of the present invention can be expressed in a mammalian cell. A typical mammalian expression vector contains a promoter element, which mediates the initiation of transcription of mRNA, a protein coding sequence, and signals required for the termination of transcription and polyadenylation of the transcript. Additional elements include enhancers, Kozak sequences and intervening sequences flanked by donor and acceptor sites for RNA splicing. Highly efficient transcription is achieved with the early and late promoters from SV40, the long terminal repeats (LTRs) from

Retroviruses, e.g., RSV, HTLVI, HIVI and the early promoter of the cytomegalovirus (CMV). However, cellular elements can also be used (e.g., the human actin promoter).

Suitable expression vectors for use in practicing the present invention include, for example, vectors such as pSVL and pMSG (Pharmacia, Uppsala, Sweden), pRSVcat (ATCC 37152), pSV2dhfr (ATCC 37146), pBC12MI (ATCC 67109), pCMVSport 2.0, and pCMVSport 3.0. Mammalian host cells that could be used include, human Hela, 293, H9 and Jurkat cells, mouse NIH3T3 and C127 cells, Cos 1, Cos 7 and CV1, quail QC1-3 cells, mouse L cells and Chinese hamster ovary (CHO) cells.

Alternatively, the polypeptide can be expressed in stable cell lines containing the polynucleotide integrated into a chromosome. The co-transfection with a selectable marker such as dhfr, gpt, neomycin, hygromycin allows the identification and isolation of the transfected cells.

The transfected gene can also be amplified to express large amounts of the encoded protein. The DHFR (dihydrofolate reductase) marker is useful in developing cell lines that carry several hundred or even several thousand copies of the gene of interest. (See, e.g., Alt, F. W., et al., J. Biol. Chem. 253:1357-1370 (1978); Hamlin, J. L. and Ma, C., Biochem. et Biophys. Acta, 1097:107-143 (1990); Page, M. J. and Sydenham, M. A., Biotechnology 9:64-68 (1991).) Another useful selection marker is the enzyme glutamine synthase (GS) (Murphy et al., Biochem J. 227:277-279 (1991); Bebbington et al., Bio/Technology 10:169-175 (1992). Using these markers, the mammalian cells are grown in selective medium and the cells with the highest resistance are selected. These cell lines contain the amplified gene(s) integrated into a chromosome. Chinese hamster ovary (CHO) and NSO cells are often used for the production of proteins.

Derivatives of the plasmid pSV2-dhfr (ATCC Accession No. 37146), the expression vectors pC4 (ATCC Accession No. 209646) and pC6 (ATCC Accession No.209647) contain the strong promoter (LTR) of the Rous Sarcoma Virus (Cullen et al., Molecular and Cellular Biology, 438-447 (March, 1985)) plus a fragment of the CMV-enhancer (Boshart et al., Cell 41:521-530 (1985).) Multiple cloning sites, e.g., with the restriction enzyme cleavage sites BamHI, XbaI and Asp718, facilitate the cloning of the gene of interest. The vectors also contain the 3' intron, the polyadenylation and termination signal of the rat preproinsulin gene, and the mouse DHFR gene under control of the SV40 early promoter.

Specifically, the plasmid pC6, for example, is digested with appropriate restriction enzymes and then dephosphorylated using calf intestinal phosphates by procedures known in the art. The vector is then isolated from a 1% agarose gel.

A polynucleotide of the present invention is amplified according to the protocol outlined in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the vector does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("Geneclean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

The amplified fragment is then digested with the same restriction enzyme and purified on a 1% agarose gel. The isolated fragment and the dephosphorylated vector are then ligated with T4 DNA ligase. *E. coli* HB101 or XL-1 Blue cells are then transformed and bacteria are identified that contain the fragment inserted into plasmid pC6 using, for instance, restriction enzyme analysis.

Chinese hamster ovary cells lacking an active DHFR gene is used for transfection. Five µg of the expression plasmid pC6 is cotransfected with 0.5 µg of the plasmid pSVneo using lipofectin (Felgner et al., *supra*). The plasmid pSV2-neo contains a dominant selectable marker, the *neo* gene from Tn5 encoding an enzyme that confers resistance to a group of antibiotics including G418. The cells are seeded in alpha minus MEM supplemented with 1 mg/ml G418. After 2 days, the cells are trypsinized and seeded in hybridoma cloning plates (Greiner, Germany) in alpha minus MEM supplemented with 10, 25, or 50 ng/ml of metothrexate plus 1 mg/ml G418. After about 10-14 days single clones are trypsinized and then seeded in 6-well petri dishes or 10 ml flasks using different concentrations of methotrexate (50 nM, 100 nM, 200 nM, 400 nM, 800 nM). Clones growing at the highest concentrations of methotrexate are then transferred to new 6-well plates containing even higher concentrations of methotrexate (1 µM, 2 µM, 5 µM, 10 mM, 20 mM). The same procedure is repeated until clones are obtained which grow at a concentration of 100 - 200 µM. Expression of the desired gene product is analyzed, for instance, by SDS-PAGE and Western blot or by reversed phase HPLC analysis.

**<u>Example 9: Protein Fusions</u>**

The polypeptides of the present invention are preferably fused to other proteins. These fusion proteins can be used for a variety of applications. For example, fusion of the present polypeptides to His-tag, HA-tag, protein A, IgG domains, and maltose binding protein facilitates purification. (See Example 5; see also EP A 394,827; Traunecker, et al., Nature 331:84-86 (1988).) Similarly, fusion to IgG-1, IgG-3, and albumin increases the halflife time in vivo. Nuclear localization signals fused to the

polypeptides of the present invention can target the protein to a specific subcellular localization, while covalent heterodimer or homodimers can increase or decrease the activity of a fusion protein. Fusion proteins can also create chimeric molecules having more than one function. Finally, fusion proteins can increase solubility and/or stability of the fused protein compared to the non-fused protein. All of the types of fusion proteins described above can be made by modifying the following protocol, which outlines the fusion of a polypeptide to an IgG molecule, or the protocol described in Example 5.

Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector.

For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 1, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced.

If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

Human IgG Fc region:
GGGATCCGGAGCCCAAATCTTCTGACAAAACTCACACATGCCCACCGTGCC
CAGCACCTGAATTCGAGGGTGCACCGTCAGTCTTCCTCTTCCCCCCAAAACC
CAAGGACACCCTCATGATCTCCCGGACTCCTGAGGTCACATGCGTGGTGGT
GGACGTAAGCCACGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGGACG
GCGTGGAGGTGCATAATGCCAAGACAAAGCCGCGGGAGGAGCAGTACAAC
AGCACGTACCGTGTGGTCAGCGTCCTCACCGTCCTGCACCAGGACTGGCTG
AATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCCCTCCCAACCCCC
ATCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCCGAGAACCACAGGT
GTACACCCTGCCCCCATCCCGGGATGAGCTGACCAAGAACCAGGTCAGCCT
GACCTGCCTGGTCAAAGGCTTCTATCCAAGCGACATCGCCGTGGAGTGGGA
GAGCAATGGGCAGCCGGAGAACAACTACAAGACCACGCCTCCCGTGCTGG
ACTCCGACGGCTCCTTCTTCCTCTACAGCAAGCTCACCGTGGACAAGAGCA

GGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGTGATGCATGAGGCTCTGC
ACAACCACTACACGCAGAAGAGCCTCTCCCTGTCTCCGGGTAAATGAGTGC
GACGGCCGCGACTCTAGAGGAT (SEQ ID NO:1)

**Example 10: Production of an Antibody from a Polypeptide**

The antibodies of the present invention can be prepared by a variety of methods. (See, Current Protocols, Chapter 2.) For example, cells expressing a polypeptide of the present invention is administered to an animal to induce the production of sera containing polyclonal antibodies. In a preferred method, a preparation of the secreted protein is prepared and purified to render it substantially free of natural contaminants. Such a preparation is then introduced into an animal in order to produce polyclonal antisera of greater specific activity.

In the most preferred method, the antibodies of the present invention are monoclonal antibodies (or protein binding fragments thereof). Such monoclonal antibodies can be prepared using hybridoma technology. (Köhler et al., Nature 256:495 (1975); Köhler et al., Eur. J. Immunol. 6:511 (1976); Köhler et al., Eur. J. Immunol. 6:292 (1976); Hammerling et al., in: Monoclonal Antibodies and T-Cell Hybridomas, Elsevier, N.Y., pp. 563-681 (1981).) In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100 μg/ml of streptomycin.

The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP2O), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands et al. (Gastroenterology 80:225-232 (1981).) The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide.

Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with

this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein-specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies.

It will be appreciated that Fab and F(ab')2 and other fragments of the antibodies of the present invention may be used according to the methods disclosed herein. Such fragments are typically produced by proteolytic cleavage, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab')2 fragments). Alternatively, secreted protein-binding fragments can be produced through the application of recombinant DNA technology or through synthetic chemistry.

For in vivo use of antibodies in humans, it may be preferable to use "humanized" chimeric monoclonal antibodies. Such antibodies can be produced using genetic constructs derived from hybridoma cells producing the monoclonal antibodies described above. Methods for producing chimeric antibodies are known in the art. (See, for review, Morrison, Science 229:1202 (1985); Oi et al., BioTechniques 4:214 (1986); Cabilly et al., U.S. Patent No. 4,816,567; Taniguchi et al., EP 171496; Morrison et al., EP 173494; Neuberger et al., WO 8601533; Robinson et al., WO 8702671; Boulianne et al., Nature 312:643 (1984); Neuberger et al., Nature 314:268 (1985).)

**Example 11: Production Of Secreted Protein For High-Throughput Screening Assays**

The following protocol produces a supernatant containing a polypeptide to be tested. This supernatant can then be used in the Screening Assays described in Examples 13-20.

First, dilute Poly-D-Lysine (644 587 Boehringer-Mannheim) stock solution (1mg/ml in PBS) 1:20 in PBS (w/o calcium or magnesium 17-516F Biowhittaker) for a working solution of 50ug/ml. Add 200 ul of this solution to each well (24 well plates) and incubate at RT for 20 minutes. Be sure to distribute the solution over each well (note: a 12-channel pipetter may be used with tips on every other channel). Aspirate off the Poly-D-Lysine solution and rinse with 1ml PBS (Phosphate Buffered Saline). The PBS should remain in the well until just prior to plating the cells and plates may be poly-lysine coated in advance for up to two weeks.

Plate 293T cells (do not carry cells past P+20) at 2 x $10^5$ cells/well in .5ml DMEM(Dulbecco's Modified Eagle Medium)(with 4.5 G/L glucose and L-glutamine (12-604F Biowhittaker))/10% heat inactivated FBS(14-503F Biowhittaker)/1x Penstrep(17-602E Biowhittaker). Let the cells grow overnight.

The next day, mix together in a sterile solution basin: 300 ul Lipofectamine (18324-012 Gibco/BRL) and 5ml Optimem I (31985070 Gibco/BRL)/96-well plate. With a small volume multi-channel pipetter, aliquot approximately 2ug of an expression vector containing a polynucleotide insert, produced by the methods described in Examples 8 or 9, into an appropriately labeled 96-well round bottom plate. With a multi-channel pipetter, add 50ul of the Lipofectamine/Optimem I mixture to each well. Pipette up and down gently to mix. Incubate at RT 15-45 minutes. After about 20 minutes, use a multi-channel pipetter to add 150ul Optimem I to each well. As a control, one plate of vector DNA lacking an insert should be transfected with each set of transfections.

Preferably, the transfection should be performed by tag-teaming the following tasks. By tag-teaming, hands on time is cut in half, and the cells do not spend too much time on PBS. First, person A aspirates off the media from four 24-well plates of cells, and then person B rinses each well with .5-1ml PBS. Person A then aspirates off PBS rinse, and person B, using a12-channel pipetter with tips on every other channel, adds the 200ul of DNA/Lipofectamine/Optimem I complex to the odd wells first, then to the even wells, to each row on the 24-well plates. Incubate at 37°C for 6 hours.

While cells are incubating, prepare appropriate media, either 1%BSA in DMEM with 1x penstrep, or CHO-5 media (116.6 mg/L of CaCl2 (anhyd); 0.00130 mg/L $CuSO_4$-5$H_2O$; 0.050 mg/L of $Fe(NO_3)_3$-9$H_2O$; 0.417 mg/L of $FeSO_4$-7$H_2O$; 311.80 mg/L of Kcl; 28.64 mg/L of $MgCl_2$; 48.84 mg/L of $MgSO_4$; 6995.50 mg/L of NaCl; 2400.0 mg/L of $NaHCO_3$; 62.50 mg/L of $NaH_2PO_4$-$H_20$; 71.02 mg/L of $Na_2$HPO4; .4320 mg/L of $ZnSO_4$-7$H_2O$; .002 mg/L of Arachidonic Acid ; 1.022 mg/L of Cholesterol; .070 mg/L of DL-alpha-Tocopherol-Acetate; 0.0520 mg/L of Linoleic Acid; 0.010 mg/L of Linolenic Acid; 0.010 mg/L of Myristic Acid; 0.010 mg/L of Oleic Acid; 0.010 mg/L of Palmitric Acid; 0.010 mg/L of Palmitic Acid; 100 mg/L of Pluronic F-68; 0.010 mg/L of Stearic Acid; 2.20 mg/L of Tween 80; 4551 mg/L of D-Glucose; 130.85 mg/ml of L- Alanine; 147.50 mg/ml of L-Arginine-HCL; 7.50 mg/ml of L-Asparagine-$H_20$; 6.65 mg/ml of L-Aspartic Acid; 29.56 mg/ml of L-Cystine-2HCL-$H_20$; 31.29 mg/ml of L-Cystine-2HCL; 7.35 mg/ml of L-Glutamic Acid; 365.0 mg/ml of L-Glutamine; 18.75 mg/ml of Glycine; 52.48 mg/ml of L-Histidine-HCL-$H_20$; 106.97 mg/ml of L-Isoleucine; 111.45 mg/ml of L-Leucine; 163.75 mg/ml of L-

Lysine HCL; 32.34 mg/ml of L-Methionine; 68.48 mg/ml of L-Phenylalainine; 40.0 mg/ml of L-Proline; 26.25 mg/ml of L-Serine; 101.05 mg/ml of L-Threonine; 19.22 mg/ml of L-Tryptophan; 91.79 mg/ml of L-Tryrosine-2Na-$2H_20$; 99.65 mg/ml of L-Valine; 0.0035 mg/L of Biotin; 3.24 mg/L of D-Ca Pantothenate; 11.78 mg/L of Choline Chloride; 4.65 mg/L of Folic Acid; 15.60 mg/L of i-Inositol; 3.02 mg/L of Niacinamide; 3.00 mg/L of Pyridoxal HCL; 0.031 mg/L of Pyridoxine HCL; 0.319 mg/L of Riboflavin; 3.17 mg/L of Thiamine HCL; 0.365 mg/L of Thymidine; and 0.680 mg/L of Vitamin $B_{12}$; 25 mM of HEPES Buffer; 2.39 mg/L of Na Hypoxanthine; 0.105 mg/L of Lipoic Acid; 0.081 mg/L of Sodium Putrescine-2HCL; 55.0 mg/L of Sodium Pyruvate; 0.0067 mg/L of Sodium Selenite; 20uM of Ethanolamine; 0.122 mg/L of Ferric Citrate; 41.70 mg/L of Methyl-B-Cyclodextrin complexed with Linoleic Acid; 33.33 mg/L of Methyl-B-Cyclodextrin complexed with Oleic Acid; and 10 mg/L of Methyl-B-Cyclodextrin complexed with Retinal) with 2mm glutamine and 1x penstrep. (BSA (81-068-3 Bayer) 100gm dissolved in 1L DMEM for a 10% BSA stock solution). Filter the media and collect 50 ul for endotoxin assay in 15ml polystyrene conical.

The transfection reaction is terminated, preferably by tag-teaming, at the end of the incubation period. Person A aspirates off the transfection media, while person B adds 1.5ml appropriate media to each well. Incubate at 37°C for 45 or 72 hours depending on the media used: 1%BSA for 45 hours or CHO-5 for 72 hours.

On day four, using a 300ul multichannel pipetter, aliquot 600ul in one 1ml deep well plate and the remaining supernatant into a 2ml deep well. The supernatants from each well can then be used in the assays described in Examples 13-20.

It is specifically understood that when activity is obtained in any of the assays described below using a supernatant, the activity originates from either the polypeptide directly (e.g., as a secreted protein) or by the polypeptide inducing expression of other proteins, which are then secreted into the supernatant. Thus, the invention further provides a method of identifying the protein in the supernatant characterized by an activity in a particular assay.

## Example 12: Construction of GAS Reporter Construct

One signal transduction pathway involved in the differentiation and proliferation of cells is called the Jaks-STATs pathway. Activated proteins in the Jaks-STATs pathway bind to gamma activation site "GAS" elements or interferon-sensitive responsive element ("ISRE"), located in the promoter of many genes. The binding of a protein to these elements alter the expression of the associated gene.

GAS and ISRE elements are recognized by a class of transcription factors called Signal Transducers and Activators of Transcription, or "STATs." There are six members of the STATs family. Stat1 and Stat3 are present in many cell types, as is Stat2 (as response to IFN-alpha is widespread). Stat4 is more restricted and is not in many cell types though it has been found in T helper class I, cells after treatment with IL-12. Stat5 was originally called mammary growth factor, but has been found at higher concentrations in other cells including myeloid cells. It can be activated in tissue culture cells by many cytokines.

The STATs are activated to translocate from the cytoplasm to the nucleus upon tyrosine phosphorylation by a set of kinases known as the Janus Kinase ("Jaks") family. Jaks represent a distinct family of soluble tyrosine kinases and include Tyk2, Jak1, Jak2, and Jak3. These kinases display significant sequence similarity and are generally catalytically inactive in resting cells.

The Jaks are activated by a wide range of receptors summarized in the Table below. (Adapted from review by Schidler and Darnell, Ann. Rev. Biochem. 64:621-51 (1995).) A cytokine receptor family, capable of activating Jaks, is divided into two groups: (a) Class 1 includes receptors for IL-2, IL-3, IL-4, IL-6, IL-7, IL-9, IL-11, IL-12, IL-15, Epo, PRL, GH, G-CSF, GM-CSF, LIF, CNTF, and thrombopoietin; and (b) Class 2 includes IFN-a, IFN-g, and IL-10. The Class 1 receptors share a conserved cysteine motif (a set of four conserved cysteines and one tryptophan) and a WSXWS motif (a membrane proxial region encoding Trp-Ser-Xxx-Trp-Ser (SEQ ID NO:2)).

Thus, on binding of a ligand to a receptor, Jaks are activated, which in turn activate STATs, which then translocate and bind to GAS elements. This entire process is encompassed in the Jaks-STATs signal transduction pathway.

Therefore, activation of the Jaks-STATs pathway, reflected by the binding of the GAS or the ISRE element, can be used to indicate proteins involved in the proliferation and differentiation of cells. For example, growth factors and cytokines are known to activate the Jaks-STATs pathway. (See Table below.) Thus, by using GAS elements linked to reporter molecules, activators of the Jaks-STATs pathway can be identified.

| Ligand | tyk2 | JAKs Jak1 | Jak2 | Jak3 | STATS | GAS(elements) or ISRE |
|---|---|---|---|---|---|---|
| IFN family | | | | | | |
| IFN-a/B | + | + | - | - | 1,2,3 | ISRE |
| IFN-g | | + | + | - | 1 | GAS (IRF1>Lys6>IFP) |
| Il-10 | + | ? | ? | - | 1,3 | |
| gp130 family | | | | | | |
| IL-6 (Pleiotrohic) | + | + | + | ? | 1,3 | GAS (IRF1>Lys6>IFP) |
| Il-11(Pleiotrohic) | ? | + | ? | ? | 1,3 | |
| OnM(Pleiotrohic) | ? | + | + | ? | 1,3 | |
| LIF(Pleiotrohic) | ? | + | + | ? | 1,3 | |
| CNTF(Pleiotrohic) | -/+ | + | + | ? | 1,3 | |
| G-CSF(Pleiotrohic) | ? | + | ? | ? | 1,3 | |
| IL-12(Pleiotrohic) | + | - | + | + | 1,3 | |
| g-C family | | | | | | |
| IL-2 (lymphocytes) | - | + | - | + | 1,3,5 | GAS |
| IL-4 (lymph/myeloid) | - | + | - | + | 6 | GAS (IRF1 = IFP >>Ly6)(IgH) |
| IL-7 (lymphocytes) | - | + | - | + | 5 | GAS |
| IL-9 (lymphocytes) | - | + | - | + | 5 | GAS |
| IL-13 (lymphocyte) | - | + | ? | ? | 6 | GAS |
| IL-15 | ? | + | ? | + | 5 | GAS |
| gp140 family | | | | | | |
| IL-3 (myeloid) | - | - | + | - | 5 | GAS (IRF1>IFP>>Ly6) |
| IL-5 (myeloid) | - | - | + | - | 5 | GAS |
| GM-CSF (myeloid) | - | - | + | - | 5 | GAS |
| Growth hormone family | | | | | | |
| GH | ? | - | + | - | 5 | |
| PRL | ? | +/- | + | - | 1,3,5 | |
| EPO | ? | - | + | - | 5 | GAS(B-CAS>IRF1=IFP>>Ly6) |
| Receptor Tyrosine Kinases | | | | | | |
| EGF | ? | + | + | - | 1,3 | GAS (IRF1) |
| PDGF | ? | + | + | - | 1,3 | |
| CSF-1 | ? | + | + | - | 1,3 | GAS (not IRF1) |

To construct a synthetic GAS containing promoter element, which is used in the Biological Assays described in Examples 13-14, a PCR based strategy is employed to generate a GAS-SV40 promoter sequence. The 5' primer contains four tandem copies of the GAS binding site found in the IRF1 promoter and previously demonstrated to bind STATs upon induction with a range of cytokines (Rothman et al., Immunity 1:457-468 (1994).), although other GAS or ISRE elements can be used instead. The 5' primer also contains 18bp of sequence complementary to the SV40 early promoter sequence and is flanked with an XhoI site. The sequence of the 5' primer is:
5':GCGCCTCGAGATTTCCCCGAAATCTAGATTTCCCCGAAATGATTTCCCCG AAATGATTTCCCCGAAATATCTGCCATCTCAATTAG:3' (SEQ ID NO:3)

The downstream primer is complementary to the SV40 promoter and is flanked with a Hind III site: 5':GCGGCAAGCTTTTTGCAAAGCCTAGGC:3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the B-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI/Hind III and subcloned into BLSK2-. (Stratagene.) Sequencing with forward and reverse primers confirms that the insert contains the following sequence:
5':<u>CTCGAG</u>ATTTCCCCGAAATCTAGATTTCCCCGAAATGATTTCCCCGAAATG ATTTCCCCGAAATATCTGCCATCTCAATTAGTCAGCAACCATAGTCCCGCCC CTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGC CCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGC CTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTT TGCAAA<u>AAGCTT</u>:3' (SEQ ID NO:5)

With this GAS promoter element linked to the SV40 promoter, a GAS:SEAP2 reporter construct is next engineered. Here, the reporter molecule is a secreted alkaline phosphatase, or "SEAP." Clearly, however, any reporter molecule can be instead of SEAP, in this or in any of the other Examples. Well known reporter molecules that can be used instead of SEAP include chloramphenicol acetyltransferase (CAT), luciferase, alkaline phosphatase, B-galactosidase, green fluorescent protein (GFP), or any protein detectable by an antibody.

The above sequence confirmed synthetic GAS-SV40 promoter element is subcloned into the pSEAP-Promoter vector obtained from Clontech using HindIII and XhoI, effectively replacing the SV40 promoter with the amplified GAS:SV40 promoter element, to create the GAS-SEAP vector. However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

Thus, in order to generate mammalian stable cell lines expressing the GAS-SEAP reporter, the GAS-SEAP cassette is removed from the GAS-SEAP vector using SalI and NotI, and inserted into a backbone vector containing the neomycin resistance gene, such as pGFP-1 (Clontech), using these restriction sites in the multiple cloning site, to create the GAS-SEAP/Neo vector. Once this vector is transfected into mammalian cells, this vector can then be used as a reporter molecule for GAS binding as described in Examples 13-14.

Other constructs can be made using the above description and replacing GAS with a different promoter sequence. For example, construction of reporter molecules containing NFK-B and EGR promoter sequences are described in Examples 15 and 16. However, many other promoters can be substituted using the protocols described in these Examples. For instance, SRE, IL-2, NFAT, or Osteocalcin promoters can be substituted, alone or in combination (e.g., GAS/NF-KB/EGR, GAS/NF-KB, Il-2/NFAT, or NF-KB/GAS). Similarly, other cell lines can be used to test reporter construct activity, such as HELA (epithelial), HUVEC (endothelial), Reh (B-cell), Saos-2 (osteoblast), HUVAC (aortic), or Cardiomyocyte.

**Example 13: High-Throughput Screening Assay for T-cell Activity.**

The following protocol is used to assess T-cell activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate T-cells. T-cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The T-cell used in this assay is Jurkat T-cells (ATCC Accession No. TIB-152), although Molt-3 cells (ATCC Accession No. CRL-1552) and Molt-4 cells (ATCC Accession No. CRL-1582) cells can also be used.

Jurkat T-cells are lymphoblastic CD4+ Th1 helper cells. In order to generate stable cell lines, approximately 2 million Jurkat cells are transfected with the GAS-SEAP/neo vector using DMRIE-C (Life Technologies)(transfection procedure described below). The transfected cells are seeded to a density of approximately 20,000 cells per well and transfectants resistant to 1 mg/ml genticin selected. Resistant colonies are expanded and then tested for their response to increasing concentrations of interferon gamma. The dose response of a selected clone is demonstrated.

Specifically, the following protocol will yield sufficient cells for 75 wells containing 200 ul of cells. Thus, it is either scaled up, or performed in multiple to generate sufficient cells for multiple 96 well plates. Jurkat cells are maintained in RPMI + 10% serum with 1%Pen-Strep. Combine 2.5 mls of OPTI-MEM (Life Technologies)

with 10 ug of plasmid DNA in a T25 flask. Add 2.5 ml OPTI-MEM containing 50 ul of DMRIE-C and incubate at room temperature for 15-45 mins.

During the incubation period, count cell concentration, spin down the required number of cells ($10^7$ per transfection), and resuspend in OPTI-MEM to a final concentration of $10^7$ cells/ml. Then add 1ml of 1 x $10^7$ cells in OPTI-MEM to T25 flask and incubate at 37°C for 6 hrs. After the incubation, add 10 ml of RPMI + 15% serum.

The Jurkat:GAS-SEAP stable reporter lines are maintained in RPMI + 10% serum, 1 mg/ml Genticin, and 1% Pen-Strep. These cells are treated with supernatants containing a polypeptide as produced by the protocol described in Example 11.

On the day of treatment with the supernatant, the cells should be washed and resuspended in fresh RPMI + 10% serum to a density of 500,000 cells per ml. The exact number of cells required will depend on the number of supernatants being screened. For one 96 well plate, approximately 10 million cells (for 10 plates, 100 million cells) are required.

Transfer the cells to a triangular reservoir boat, in order to dispense the cells into a 96 well dish, using a 12 channel pipette. Using a 12 channel pipette, transfer 200 ul of cells into each well (therefore adding 100, 000 cells per well).

After all the plates have been seeded, 50 ul of the supernatants are transferred directly from the 96 well plate containing the supernatants into each well using a 12 channel pipette. In addition, a dose of exogenous interferon gamma (0.1, 1.0, 10 ng) is added to wells H9, H10, and H11 to serve as additional positive controls for the assay.

The 96 well dishes containing Jurkat cells treated with supernatants are placed in an incubator for 48 hrs (note: this time is variable between 48-72 hrs). 35 ul samples from each well are then transferred to an opaque 96 well plate using a 12 channel pipette. The opaque plates should be covered (using sellophene covers) and stored at -20°C until SEAP assays are performed according to Example 17. The plates containing the remaining treated cells are placed at 4°C and serve as a source of material for repeating the assay on a specific well if desired.

As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate Jurkat T cells. Over 30 fold induction is typically observed in the positive control wells.

**Example 14: High-Throughput Screening Assay Identifying Myeloid Activity**

The following protocol is used to assess myeloid activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate myeloid cells. Myeloid cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The myeloid cell used in this assay is U937, a pre-monocyte cell line, although TF-1, HL60, or KG1 can be used.

To transiently transfect U937 cells with the GAS/SEAP/Neo construct produced in Example 12, a DEAE-Dextran method (Kharbanda et. al., 1994, Cell Growth & Differentiation, 5:259-265) is used. First, harvest 2x10e$^{7}$ U937 cells and wash with PBS. The U937 cells are usually grown in RPMI 1640 medium containing 10% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 mg/ml streptomycin.

Next, suspend the cells in 1 ml of 20 mM Tris-HCl (pH 7.4) buffer containing 0.5 mg/ml DEAE-Dextran, 8 ug GAS-SEAP2 plasmid DNA, 140 mM NaCl, 5 mM KCl, 375 uM $Na_2HPO_4.7H_2O$, 1 mM $MgCl_2$, and 675 uM $CaCl_2$. Incubate at 37°C for 45 min.

Wash the cells with RPMI 1640 medium containing 10% FBS and then resuspend in 10 ml complete medium and incubate at 37°C for 36 hr.

The GAS-SEAP/U937 stable cells are obtained by growing the cells in 400 ug/ml G418. The G418-free medium is used for routine growth but every one to two months, the cells should be re-grown in 400 ug/ml G418 for couple of passages.

These cells are tested by harvesting $1x10^8$ cells (this is enough for ten 96-well plates assay) and wash with PBS. Suspend the cells in 200 ml above described growth medium, with a final density of $5x10^5$ cells/ml. Plate 200 ul cells per well in the 96-well plate (or $1x10^5$ cells/well).

Add 50 ul of the supernatant prepared by the protocol described in Example 11. Incubate at 37°C for 48 to 72 hr. As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate U937 cells. Over 30 fold induction is typically observed in the positive control wells. SEAP assay the supernatant according to the protocol described in Example 17.

**Example 15: High-Throughput Screening Assay Identifying Neuronal Activity.**

When cells undergo differentiation and proliferation, a group of genes are activated through many different signal transduction pathways. One of these genes, EGR1 (early growth response gene 1), is induced in various tissues and cell types upon activation. The promoter of EGR1 is responsible for such induction. Using the EGR1 promoter linked to reporter molecules, activation of cells can be assessed.

Particularly, the following protocol is used to assess neuronal activity in PC12 cell lines. PC12 cells (rat phenochromocytoma cells) are known to proliferate and/or differentiate by activation with a number of mitogens, such as TPA (tetradecanoyl phorbol acetate), NGF (nerve growth factor), and EGF (epidermal growth factor). The EGR1 gene expression is activated during this treatment. Thus, by stably transfecting PC12 cells with a construct containing an EGR promoter linked to SEAP reporter, activation of PC12 cells can be assessed.

The EGR/SEAP reporter construct can be assembled by the following protocol. The EGR-1 promoter sequence (-633 to +1)(Sakamoto K et al., Oncogene 6:867-871 (1991)) can be PCR amplified from human genomic DNA using the following primers:

5' GCGCTCGAGGGATGACAGCGATAGAACCCCGG -3' (SEQ ID NO:6)

5' GCGAAGCTTCGCGACTCCCCGGATCCGCCTC-3' (SEQ ID NO:7)

Using the GAS:SEAP/Neo vector produced in Example 12, EGR1 amplified product can then be inserted into this vector. Linearize the GAS:SEAP/Neo vector using restriction enzymes XhoI/HindIII, removing the GAS/SV40 stuffer. Restrict the EGR1 amplified product with these same enzymes. Ligate the vector and the EGR1 promoter.

To prepare 96 well-plates for cell culture, two mls of a coating solution (1:30 dilution of collagen type I (Upstate Biotech Inc. Cat#08-115) in 30% ethanol (filter sterilized)) is added per one 10 cm plate or 50 ml per well of the 96-well plate, and allowed to air dry for 2 hr.

PC12 cells are routinely grown in RPMI-1640 medium (Bio Whittaker) containing 10% horse serum (JRH BIOSCIENCES, Cat. # 12449-78P), 5% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 ug/ml streptomycin on a precoated 10 cm tissue culture dish. One to four split is done every three to four days. Cells are removed from the plates by scraping and resuspended with pipetting up and down for more than 15 times.

Transfect the EGR/SEAP/Neo construct into PC12 using the Lipofectamine protocol described in Example 11. EGR-SEAP/PC12 stable cells are obtained by growing the cells in 300 ug/ml G418. The G418-free medium is used for routine

growth but every one to two months, the cells should be re-grown in 300 ug/ml G418 for couple of passages.

To assay for neuronal activity, a 10 cm plate with cells around 70 to 80% confluent is screened by removing the old medium. Wash the cells once with PBS (Phosphate buffered saline). Then starve the cells in low serum medium (RPMI-1640 containing 1% horse serum and 0.5% FBS with antibiotics) overnight.

The next morning, remove the medium and wash the cells with PBS. Scrape off the cells from the plate, suspend the cells well in 2 ml low serum medium. Count the cell number and add more low serum medium to reach final cell density as $5x10^5$ cells/ml.

Add 200 ul of the cell suspension to each well of 96-well plate (equivalent to $1x10^5$ cells/well). Add 50 ul supernatant produced by Example 11, 37°C for 48 to 72 hr. As a positive control, a growth factor known to activate PC12 cells through EGR can be used, such as 50 ng/ul of Neuronal Growth Factor (NGF). Over fifty-fold induction of SEAP is typically seen in the positive control wells. SEAP assay the supernatant according to Example 17.

**Example 16: High-Throughput Screening Assay for T-cell Activity**

NF-κB (Nuclear Factor κB) is a transcription factor activated by a wide variety of agents including the inflammatory cytokines IL-1 and TNF, CD30 and CD40, lymphotoxin-alpha and lymphotoxin-beta, by exposure to LPS or thrombin, and by expression of certain viral gene products. As a transcription factor, NF-κB regulates the expression of genes involved in immune cell activation, control of apoptosis (NF-κB appears to shield cells from apoptosis), B and T-cell development, anti-viral and antimicrobial responses, and multiple stress responses.

In non-stimulated conditions, NF- κB is retained in the cytoplasm with I-κB (Inhibitor κB). However, upon stimulation, I- κB is phosphorylated and degraded, causing NF- κB to shuttle to the nucleus, thereby activating transcription of target genes. Target genes activated by NF- κB include IL-2, IL-6, GM-CSF, ICAM-1 and class 1 MHC.

Due to its central role and ability to respond to a range of stimuli, reporter constructs utilizing the NF-κB promoter element are used to screen the supernatants produced in Example 11. Activators or inhibitors of NF-kB would be useful in treating

diseases. For example, inhibitors of NF-κB could be used to treat those diseases related to the acute or chronic activation of NF-kB, such as rheumatoid arthritis.

To construct a vector containing the NF-κB promoter element, a PCR based strategy is employed. The upstream primer contains four tandem copies of the NF-κB binding site (GGGGACTTTCCC) (SEQ ID NO:8), 18 bp of sequence complementary to the 5' end of the SV40 early promoter sequence, and is flanked with an XhoI site:
5':GCGGCCTCGAGGGGACTTTCCCGGGGACTTTCCGGGGACTTTCCGGGAC
TTTCCATCCTGCCATCTCAATTAG:3' (SEQ ID NO:9)

The downstream primer is complementary to the 3' end of the SV40 promoter and is flanked with a Hind III site:
5':GCGGCAAGCTTTTTGCAAAGCCTAGGC:3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the pB-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI and Hind III and subcloned into BLSK2-. (Stratagene) Sequencing with the T7 and T3 primers confirms the insert contains the following sequence:

5':CTCGAGGGGACTTTCCCGGGGACTTTCCGGGGACTTTCCGGGACTTTCC
ATCTGCCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCA
TCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACT
AATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTC
CAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTT:
3' (SEQ ID NO:10)

Next, replace the SV40 minimal promoter element present in the pSEAP2-promoter plasmid (Clontech) with this NF-κB/SV40 fragment using XhoI and HindIII. However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

In order to generate stable mammalian cell lines, the NF-κB/SV40/SEAP cassette is removed from the above NF-κB/SEAP vector using restriction enzymes SalI and NotI, and inserted into a vector containing neomycin resistance. Particularly, the NF-κB/SV40/SEAP cassette was inserted into pGFP-1 (Clontech), replacing the GFP gene, after restricting pGFP-1 with SalI and NotI.

Once NF-κB/SV40/SEAP/Neo vector is created, stable Jurkat T-cells are created and maintained according to the protocol described in Example 13. Similarly, the method for assaying supernatants with these stable Jurkat T-cells is also described in Example 13. As a positive control, exogenous TNF alpha (0.1,1, 10 ng) is added to wells H9, H10, and H11, with a 5-10 fold activation typically observed.

**Example 17: Assay for SEAP Activity**

As a reporter molecule for the assays described in Examples 13-16, SEAP activity is assayed using the Tropix Phospho-light Kit (Cat. BP-400) according to the following general procedure. The Tropix Phospho-light Kit supplies the Dilution, Assay, and Reaction Buffers used below.

Prime a dispenser with the 2.5x Dilution Buffer and dispense 15 μl of 2.5x dilution buffer into Optiplates containing 35 μl of a supernatant. Seal the plates with a plastic sealer and incubate at 65°C for 30 min. Separate the Optiplates to avoid uneven heating.

Cool the samples to room temperature for 15 minutes. Empty the dispenser and prime with the Assay Buffer. Add 50 μl Assay Buffer and incubate at room temperature 5 min. Empty the dispenser and prime with the Reaction Buffer (see the table below). Add 50 μl Reaction Buffer and incubate at room temperature for 20 minutes. Since the intensity of the chemiluminescent signal is time dependent, and it takes about 10 minutes to read 5 plates on luminometer, one should treat 5 plates at each time and start the second set 10 minutes later.

Read the relative light unit in the luminometer. Set H12 as blank, and print the results. An increase in chemiluminescence indicates reporter activity.

**Reaction Buffer Formulation:**

| # of plates | Rxn buffer diluent (ml) | CSPD (ml) |
|---|---|---|
| 10 | 60 | 3 |
| 11 | 65 | 3.25 |
| 12 | 70 | 3.5 |
| 13 | 75 | 3.75 |
| 14 | 80 | 4 |
| 15 | 85 | 4.25 |
| 16 | 90 | 4.5 |
| 17 | 95 | 4.75 |
| 18 | 100 | 5 |
| 19 | 105 | 5.25 |
| 20 | 110 | 5.5 |
| 21 | 115 | 5.75 |
| 22 | 120 | 6 |

| | | |
|---|---|---|
| 23 | 125 | 6.25 |
| 24 | 130 | 6.5 |
| 25 | 135 | 6.75 |
| 26 | 140 | 7 |
| 27 | 145 | 7.25 |
| 28 | 150 | 7.5 |
| 29 | 155 | 7.75 |
| 30 | 160 | 8 |
| 31 | 165 | 8.25 |
| 32 | 170 | 8.5 |
| 33 | 175 | 8.75 |
| 34 | 180 | 9 |
| 35 | 185 | 9.25 |
| 36 | 190 | 9.5 |
| 37 | 195 | 9.75 |
| 38 | 200 | 10 |
| 39 | 205 | 10.25 |
| 40 | 210 | 10.5 |
| 41 | 215 | 10.75 |
| 42 | 220 | 11 |
| 43 | 225 | 11.25 |
| 44 | 230 | 11.5 |
| 45 | 235 | 11.75 |
| 46 | 240 | 12 |
| 47 | 245 | 12.25 |
| 48 | 250 | 12.5 |
| 49 | 255 | 12.75 |
| 50 | 260 | 13 |

**Example 18: High-Throughput Screening Assay Identifying Changes in Small Molecule Concentration and Membrane Permeability**

Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium, sodium, and pH, as well as alter membrane potential. These alterations can be measured in an assay to identify supernatants which bind to receptors of a particular cell. Although the following protocol describes an assay for calcium, this protocol can easily be modified to detect changes in potassium, sodium, pH, membrane potential, or any other small molecule which is detectable by a fluorescent probe.

The following assay uses Fluorometric Imaging Plate Reader ("FLIPR") to measure changes in fluorescent molecules (Molecular Probes) that bind small molecules. Clearly, any fluorescent molecule detecting a small molecule can be used instead of the calcium fluorescent molecule, fluo-3, used here.

For adherent cells, seed the cells at 10,000 -20,000 cells/well in a Co-star black 96-well plate with clear bottom. The plate is incubated in a $CO_2$ incubator for 20 hours. The adherent cells are washed two times in Biotek washer with 200 ul of HBSS (Hank's Balanced Salt Solution) leaving 100 ul of buffer after the final wash.

A stock solution of 1 mg/ml fluo-3 is made in 10% pluronic acid DMSO. To load the cells with fluo-3, 50 ul of 12 ug/ml fluo-3 is added to each well. The plate is incubated at 37°C in a $CO_2$ incubator for 60 min. The plate is washed four times in the Biotek washer with HBSS leaving 100 ul of buffer.

For non-adherent cells, the cells are spun down from culture media. Cells are re-suspended to 2-5x$10^6$ cells/ml with HBSS in a 50-ml conical tube. 4 ul of 1 mg/ml fluo-3 solution in 10% pluronic acid DMSO is added to each ml of cell suspension. The tube is then placed in a 37°C water bath for 30-60 min. The cells are washed twice with HBSS, resuspended to 1x$10^6$ cells/ml, and dispensed into a microplate, 100 ul/well. The plate is centrifuged at 1000 rpm for 5 min. The plate is then washed once in Denley CellWash with 200 ul, followed by an aspiration step to 100 ul final volume.

For a non-cell based assay, each well contains a fluorescent molecule, such as fluo-3. The supernatant is added to the well, and a change in fluorescence is detected.

To measure the fluorescence of intracellular calcium, the FLIPR is set for the following parameters: (1) System gain is 300-800 mW; (2) Exposure time is 0.4 second; (3) Camera F/stop is F/2; (4) Excitation is 488 nm; (5) Emission is 530 nm; and (6) Sample addition is 50 ul. Increased emission at 530 nm indicates an extracellular signaling event which has resulted in an increase in the intracellular $Ca^{++}$ concentration.

## Example 19: High-Throughput Screening Assay Identifying Tyrosine Kinase Activity

The Protein Tyrosine Kinases (PTK) represent a diverse group of transmembrane and cytoplasmic kinases. Within the Receptor Protein Tyrosine Kinase RPTK) group are receptors for a range of mitogenic and metabolic growth factors including the PDGF, FGF, EGF, NGF, HGF and Insulin receptor subfamilies. In addition there are a large family of RPTKs for which the corresponding ligand is unknown. Ligands for RPTKs include mainly secreted small proteins, but also membrane-bound and extracellular matrix proteins.

Activation of RPTK by ligands involves ligand-mediated receptor dimerization, resulting in transphosphorylation of the receptor subunits and activation of the cytoplasmic tyrosine kinases. The cytoplasmic tyrosine kinases include receptor associated tyrosine kinases of the src-family (e.g., src, yes, lck, lyn, fyn) and non-receptor linked and cytosolic protein tyrosine kinases, such as the Jak family, members of which mediate signal transduction triggered by the cytokine superfamily of receptors (e.g., the Interleukins, Interferons, GM-CSF, and Leptin).

Because of the wide range of known factors capable of stimulating tyrosine kinase activity, the identification of novel human secreted proteins capable of activating tyrosine kinase signal transduction pathways are of interest. Therefore, the following protocol is designed to identify those novel human secreted proteins capable of activating the tyrosine kinase signal transduction pathways.

Seed target cells (e.g., primary keratinocytes) at a density of approximately 25,000 cells per well in a 96 well Loprodyne Silent Screen Plates purchased from Nalge Nunc (Naperville, IL). The plates are sterilized with two 30 minute rinses with 100% ethanol, rinsed with water and dried overnight. Some plates are coated for 2 hr with 100 ml of cell culture grade type I collagen (50 mg/ml), gelatin (2%) or polylysine (50 mg/ml), all of which can be purchased from Sigma Chemicals (St. Louis, MO) or 10% Matrigel purchased from Becton Dickinson (Bedford,MA), or calf serum, rinsed with PBS and stored at 4°C. Cell growth on these plates is assayed by seeding 5,000 cells/well in growth medium and indirect quantitation of cell number through use of alamarBlue as described by the manufacturer Alamar Biosciences, Inc. (Sacramento, CA) after 48 hr. Falcon plate covers #3071 from Becton Dickinson (Bedford,MA) are used to cover the Loprodyne Silent Screen Plates. Falcon Microtest III cell culture plates can also be used in some proliferation experiments.

To prepare extracts, A431 cells are seeded onto the nylon membranes of Loprodyne plates (20,000/200ml/well) and cultured overnight in complete medium. Cells are quiesced by incubation in serum-free basal medium for 24 hr. After 5-20 minutes treatment with EGF (60ng/ml) or 50 ul of the supernatant produced in Example 11, the medium was removed and 100 ml of extraction buffer ((20 mM HEPES pH 7.5, 0.15 M NaCl, 1% Triton X-100, 0.1% SDS, 2 mM Na3VO4, 2 mM Na4P2O7 and a cocktail of protease inhibitors (# 1836170) obtained from Boeheringer Mannheim (Indianapolis, IN) is added to each well and the plate is shaken on a rotating shaker for 5 minutes at 4°C. The plate is then placed in a vacuum transfer manifold and the extract filtered through the 0.45 mm membrane bottoms of each well using house vacuum. Extracts are collected in a 96-well catch/assay plate in the bottom of the vacuum manifold and immediately placed on ice. To obtain extracts clarified by centrifugation, the content of each well, after detergent solubilization for 5 minutes, is removed and centrifuged for 15 minutes at 4°C at 16,000 x g.

Test the filtered extracts for levels of tyrosine kinase activity. Although many methods of detecting tyrosine kinase activity are known, one method is described here.

Generally, the tyrosine kinase activity of a supernatant is evaluated by determining its ability to phosphorylate a tyrosine residue on a specific substrate (a

biotinylated peptide). Biotinylated peptides that can be used for this purpose include PSK1 (corresponding to amino acids 6-20 of the cell division kinase cdc2-p34) and PSK2 (corresponding to amino acids 1-17 of gastrin). Both peptides are substrates for a range of tyrosine kinases and are available from Boehringer Mannheim.

The tyrosine kinase reaction is set up by adding the following components in order. First, add 10ul of 5uM Biotinylated Peptide, then 10ul ATP/$Mg_{2+}$ (5mM ATP/50mM $MgCl_2$), then 10ul of 5x Assay Buffer (40mM imidazole hydrochloride, pH7.3, 40 mM beta-glycerophosphate, 1mM EGTA, 100mM $MgCl_2$, 5 mM $MnCl_2$, 0.5 mg/ml BSA), then 5ul of Sodium Vanadate(1mM), and then 5ul of water. Mix the components gently and preincubate the reaction mix at 30°C for 2 min. Initial the reaction by adding 10ul of the control enzyme or the filtered supernatant.

The tyrosine kinase assay reaction is then terminated by adding 10 ul of 120mm EDTA and place the reactions on ice.

Tyrosine kinase activity is determined by transferring 50 ul aliquot of reaction mixture to a microtiter plate (MTP) module and incubating at 37°C for 20 min. This allows the streptavadin coated 96 well plate to associate with the biotinylated peptide. Wash the MTP module with 300ul/well of PBS four times. Next add 75 ul of anti-phospotyrosine antibody conjugated to horse radish peroxidase(anti-P-Tyr-POD(0.5u/ml)) to each well and incubate at 37°C for one hour. Wash the well as above.

Next add 100ul of peroxidase substrate solution (Boehringer Mannheim) and incubate at room temperature for at least 5 mins (up to 30 min). Measure the absorbance of the sample at 405 nm by using ELISA reader. The level of bound peroxidase activity is quantitated using an ELISA reader and reflects the level of tyrosine kinase activity.

## Example 20: High-Throughput Screening Assay Identifying Phosphorylation Activity

As a potential alternative and/or compliment to the assay of protein tyrosine kinase activity described in Example 19, an assay which detects activation (phosphorylation) of major intracellular signal transduction intermediates can also be used. For example, as described below one particular assay can detect tyrosine phosphorylation of the Erk-1 and Erk-2 kinases. However, phosphorylation of other molecules, such as Raf, JNK, p38 MAP, Map kinase kinase (MEK), MEK kinase, Src, Muscle specific kinase (MuSK), IRAK, Tec, and Janus, as well as any other

phosphoserine, phosphotyrosine, or phosphothreonine molecule, can be detected by substituting these molecules for Erk-1 or Erk-2 in the following assay.

Specifically, assay plates are made by coating the wells of a 96-well ELISA plate with 0.1ml of protein G (1ug/ml) for 2 hr at room temp, (RT). The plates are then rinsed with PBS and blocked with 3% BSA/PBS for 1 hr at RT. The protein G plates are then treated with 2 commercial monoclonal antibodies (100ng/well) against Erk-1 and Erk-2 (1 hr at RT) (Santa Cruz Biotechnology). (To detect other molecules, this step can easily be modified by substituting a monoclonal antibody detecting any of the above described molecules.) After 3-5 rinses with PBS, the plates are stored at 4°C until use.

A431 cells are seeded at 20,000/well in a 96-well Loprodyne filterplate and cultured overnight in growth medium. The cells are then starved for 48 hr in basal medium (DMEM) and then treated with EGF (6ng/well) or 50 ul of the supernatants obtained in Example 11 for 5-20 minutes. The cells are then solubilized and extracts filtered directly into the assay plate.

After incubation with the extract for 1 hr at RT, the wells are again rinsed. As a positive control, a commercial preparation of MAP kinase (10ng/well) is used in place of A431 extract. Plates are then treated with a commercial polyclonal (rabbit) antibody (1ug/ml) which specifically recognizes the phosphorylated epitope of the Erk-1 and Erk-2 kinases (1 hr at RT). This antibody is biotinylated by standard procedures. The bound polyclonal antibody is then quantitated by successive incubations with Europium-streptavidin and Europium fluorescence enhancing reagent in the Wallac DELFIA instrument (time-resolved fluorescence). An increased fluorescent signal over background indicates a phosphorylation.

## Example 21: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide

RNA isolated from entire families or individual patients presenting with a phenotype of interest (such as a disease) is be isolated. cDNA is then generated from these RNA samples using protocols known in the art. (See, Sambrook.) The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO:X. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky, D., et al., Science 252:706 (1991).

PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies).

The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations is then cloned and sequenced to validate the results of the direct sequencing.

PCR products is cloned into T-tailed vectors as described in Holton, T.A. and Graham, M.W., Nucleic Acids Research, 19:1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements are also observed as a method of determining alterations in a gene corresponding to a polynucleotide. Genomic clones isolated according to Example 2 are nick-translated with digoxigenindeoxy-uridine 5'-triphosphate (Boehringer Manheim), and FISH performed as described in Johnson, Cg. et al., Methods Cell Biol. 35:73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C- and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. (Johnson, Cv. et al., Genet. Anal. Tech. Appl., 8:75 (1991).) Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

**Example 22: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample**

A polypeptide of the present invention can be detected in a biological sample, and if an increased or decreased level of the polypeptide is detected, this polypeptide is a marker for a particular phenotype. Methods of detection are numerous, and thus, it is understood that one skilled in the art can modify the following assay to fit their particular needs.

For example, antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 ug/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described in Example 10.

The wells are blocked so that non-specific binding of the polypeptide to the well is reduced.

The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbounded polypeptide.

Next, 50 ul of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbounded conjugate.

Add 75 ul of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution to each well and incubate 1 hour at room temperature. Measure the reaction by a microtiter plate reader. Prepare a standard curve, using serial dilutions of a control sample, and plot polypeptide concentration on the X-axis (log scale) and fluorescence or absorbance of the Y-axis (linear scale). Interpolate the concentration of the polypeptide in the sample using the standard curve.

## Example 23: Formulating a Polypeptide

The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 μg/kg/day to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 μg/kg/hour to about 50 μg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracistemally, intravaginally,

intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), bucally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semi-permeable polymer matrices in the form of shaped articles, e.g., films, or mirocapsules. Sustained-release matrices include polylactides (U.S. Pat. No. 3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22:547-556 (1983)), poly (2- hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15:167-277 (1981), and R. Langer, Chem. Tech. 12:98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D- (-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE 3,218,121; Epstein et al., Proc. Natl. Acad. Sci. USA 82:3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77:4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U.S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, i.e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation. For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides.

Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e.g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrins; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e.g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampoules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1% (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

**Example 24: Method of Treating Decreased Levels of the Polypeptide**

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 ug/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided in Example 23.

**Example 25: Method of Treating Increased Levels of the Polypeptide**

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided in Example 23.

**Example 26: Method of Treatment Using Gene Therapy**

One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e.g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks.

pMV-7 (Kirschmeier, P.T. et al., DNA, 7:219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5' and 3' end sequences respectively as set forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to transform bacteria HB101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

The amphotropic pA317 or GP+am12 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media. If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

It will be clear that the invention may be practiced otherwise than as particularly described in the foregoing description and examples. Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, are within the scope of the appended claims.

The entire disclosure of each document cited (including patents, patent applications, journal articles, abstracts, laboratory manuals, books, or other disclosures) in the Background of the Invention, Detailed Description, and Examples is hereby incorporated herein by reference.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Human Genome Sciences, Inc., et al.
(ii) TITLE OF INVENTION: 32 Human Secreted Proteins
(iii) NUMBER OF SEQUENCES: 120

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Human Genome Sciences, Inc.
(B) STREET: 9410 Key West Avenue
(C) CITY: Rockville
(D) STATE: Maryland
(E) COUNTRY: USA
(F) ZIP: 20850

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
(B) COMPUTER: HP Vectra 486/33
(C) OPERATING SYSTEM: MSDOS version 6.2
(D) SOFTWARE: ASCII Text

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:
(B) FILING DATE: May 27, 1998
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER:
(B) FILING DATE:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: A. Anders Brookes
(B) REGISTRATION NUMBER: 36,373
(C) REFERENCE/DOCKET NUMBER: PZ006PCT

(vi) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (301) 309-8504
(B) TELEFAX: (301) 309-8439

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 733 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
GGGATCCGGA GCCCAAATCT TCTGACAAAA CTCACACATG CCCACCGTGC CCAGCACCTG     60
AATTCGAGGG TGCACCGTCA GTCTTCCTCT TCCCCCCAAA ACCCAAGGAC ACCCTCATGA    120
TCTCCCGGAC TCCTGAGGTC ACATGCGTGG TGGTGGACGT AAGCCACGAA GACCCTGAGG    180
TCAAGTTCAA CTGGTACGTG GACGGCGTGG AGGTGCATAA TGCCAAGACA AAGCCGCGGG    240
AGGAGCAGTA CAACAGCACG TACCGTGTGG TCAGCGTCCT CACCGTCCTG CACCAGGACT    300
GGCTGAATGG CAAGGAGTAC AAGTGCAAGG TCTCCAACAA AGCCCTCCCA ACCCCCATCG    360
AGAAAACCAT CTCCAAAGCC AAAGGGCAGC CCCGAGAACC ACAGGTGTAC ACCCTGCCCC    420
CATCCCGGGA TGAGCTGACC AAGAACCAGG TCAGCCTGAC CTGCCTGGTC AAAGGCTTCT    480
ATCCAAGCGA CATCGCCGTG GAGTGGGAGA GCAATGGGCA GCCGGAGAAC AACTACAAGA    540
CCACGCCTCC CGTGCTGGAC TCCGACGGCT CCTTCTTCCT CTACAGCAAG CTCACCGTGG    600
ACAAGAGCAG GTGGCAGCAG GGGAACGTCT TCTCATGCTC CGTGATGCAT GAGGCTCTGC    660
ACAACCACTA CACGCAGAAG AGCCTCTCCC TGTCTCCGGG TAAATGAGTG CGACGGCCGC    720
GACTCTAGAG GAT                                                       733
```

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```
Trp Ser Xaa Trp Ser
  1               5
```

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 86 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GCGCCTCGAG ATTTCCCCGA AATCTAGATT TCCCCGAAAT GATTTCCCCG AAATGATTTC 60

CCCGAAATAT CTGCCATCTC AATTAG 86

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

GCGGCAAGCT TTTTGCAAAG CCTAGGC 27

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 271 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CTCGAGATTT CCCCGAAATC TAGATTTCCC CGAAATGATT TCCCCGAAAT GATTTCCCCG 60

AAATATCTGC CATCTCAATT AGTCAGCAAC CATAGTCCCG CCCCTAACTC CGCCCATCCC 120

GCCCCTAACT CCGCCCAGTT CCGCCCATTC TCCGCCCCAT GGCTGACTAA TTTTTTTTAT 180

TTATGCAGAG GCCGAGGCCG CCTCGGCCTC TGAGCTATTC CAGAAGTAGT GAGGAGGCTT 240

TTTTGGAGGC CTAGGCTTTT GCAAAAAGCT T 271

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

GCGCTCGAGG GATGACAGCG ATAGAACCCC GG 32

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GCGAAGCTTC GCGACTCCCC GGATCCGCCT C 31

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GGGGACTTTC CC 12

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 73 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

GCGGCCTCGA GGGGACTTTC CCGGGGACTT TCCGGGGACT TTCCGGGACT TTCCATCCTG 60

CCATCTCAAT TAG 73

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 256 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

```
CTCGAGGGGA CTTTCCCGGG GACTTTCCGG GGACTTTCCG GGACTTTCCA TCTGCCATCT     60
CAATTAGTCA GCAACCATAG TCCCGCCCCT AACTCCGCCC ATCCCGCCCC TAACTCCGCC    120
CAGTTCCGCC CATTCTCCGC CCCATGGCTG ACTAATTTTT TTTATTTATG CAGAGGCCGA    180
GGCCGCCTCG GCCTCTGAGC TATTCCAGAA GTAGTGAGGA GGCTTTTTTG GAGGCCTAGG    240
CTTTTGCAAA AAGCTT                                                    256
```

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1169 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

```
GGGGCGCAAA TAGGGTCAGT GGGCCGCTTG GCGKTGTTCG TTGCGGTACC AGGTCCGCGT     60
GAGGGGTTCG GGGGTTCTGG GCAGGCACAA TGGCGTCTCG AGCAGGCCCG CGAGCGGCCG    120
RCACCGACGC AGCGAGCTTT CAGCACCGGG AGCGCGTCGC CATGCACTAC CAGATGAGTG    180
TGACCCTCAA GTATGAAATC AAGAAGCTGA TCTACGTACA TCTGGTCATA TGGCTGCTGC    240
TGGTTGCTAA GATGAGCGTG GGACACCTGA GGCTCTTGTC ACATGATCAG GTGGCCATGC    300
CCTATCAGTG GGAATACCCG TATTTGCTGA GCATTTTGCC CTCTCTCTTG GGCCTTCTCT    360
CCTTTCCCCG CAACAACATT AGCTACCTGG TGCTCTCCAT GATCAGCATG GGACTCTTTT    420
CCATCGCTCC ACTCATTTAT GGCAGCATGG AGATGTTCCC TGCTGCACAG CCTTCTACCG    480
CCATGGCAAG GCCTACCGTT TCCTCTTTGG TTTTTCTGCC GTTTCCATCA TGTACCTGGT    540
GTTGGTGTTG GCAGTGCAAG TGCATGCCTG GCAGTTGTAC TACAGCAAGA AGCTCCTAGA    600
CTCTTGGTTC ACCAGCACAC AGGAGAAGAA GCATAAATGA AGCCTCTTTG GGGTGAAGCC    660
TGGACATCCC ATCGAATGAA AGGACACTAG TACAGCGGTT CCAAAATCCC TTCTGGTGAT    720
TTTAGCAGCT GTGATGTTGG TACCTGGTGC AGACCCAGGC CAAAGTTCTG GAAAGCTCCT    780
TTTGCCATCT GCTGAGGTGG CAAAACTATA ATTTATTCCT GGTTGGCTAG AACTGGGTGA    840
CCAACAGCTA TGAAACAAAT TTCAGCTGTT TGAAGTTGAA CTTTGAGGTT TTTCTTTAAG    900
AATGAGCTTC GTCCTTGCCT CTACTCGGTC ATTCTCCCCA TTTCCATCCA TTACCCCTTA    960
GCCATTGAGA CTAAAGGAAA TAGGGAATAA ATCAAATTAC TTCATCTCTA GGTCACGGGT   1020
CAGGAAACAT TTGGGCAGCT GCTCCCTTGG CAGCTGTGGT CTCCTCTGCA AAGCATTTTA   1080
ATTAAAAACC TCAATAAAGA TGCCCTGCCC ACAAAAAAAA AAAAAAAAAA AATTCGGGGG   1140
```

GGGGCCCGGG NAACCAATTN GCCCCTANA 1169

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1310 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

AATTCGGCAC GAGGCAGCGT CGCGCGGCCC AGTTCCCTTT TCCGGTCGGC GTGGTCTTGC 60

GAGTGGAGTG TCCGCTGTGC CCGGGCCTGC ACCATGAGCG TCCCGGCCTT CATCGACATC 120

AGTGAAGAAG ATCAGGCTGC TGAGCTTCGT GCTTATCTGA AATCTAAAGG AGCTGAGATT 180

TCAGAAGAGA ACTCGGAAGG TGGACTTCAT GTTGATTTAG CTCAAATTAT TGAAGCCTGT 240

GATGTGTGTC TGAAGGAGGA TGATAAAGAT GTTGAAAGTG TGATGAACAG TGTGGTATCC 300

CTACTCTTGA TCCTGGAACC AGACAAGCAA GAAGCTTTGA TTGAAAGCCT ATGTGAAAAG 360

CTGGTCAAAT TTCGCGAAGG TGAACGCCCG TCTCTGAGAC TGCAGTTGTT AAGCAACCTT 420

TTCCACGGGA TGGATAAGAA TACTCCTGTA AGATACACAG TGTATTGCAG CCTTATTAAA 480

GTGGCAGCAT CTTGTGGGGC CATCCAGTAC ATCCCAACTG AGCTGGATCA AGTTAGAAAA 540

TGGATTTCTG ACTGGAATCT CACCACTGAA AAAAAGCACA CCCTTTTAAG ACTACTTTAT 600

GAGGCACTTG TGGATTGTAA GAAGAGTGAT GCTGCTTCAA AAGTCATGGT GGAATTGCTC 660

GGAAGTTACA CAGAGGACAA TGCTTCCCAG GCTCGAGTTG ATGCCCACAG GTGTATTGTA 720

CGAGCATTGA AAGATCCAAA TGCATTTCTT TTTGACCACC TTCTTACTTT AAAACCAGTC 780

AAGTTTTTGG AAGGCGAGCT TATTCATGAT CTTTTAACCA TTTTTGTGAG TGCTAAATTG 840

GCATCATATG TCAAGTTTTA TCAGAATAAT AAAGACTTCA TTGATTCACT TGGCCTGTTA 900

CATGAACAGA ATATGGCAAA AATGAGACTA CTTACTTTTA TGGGAATGGC AGTAGAAAAT 960

AAGGAAATTT CTTTTGACAC AATGCAGCAA GAACTTCAGA TTGGAGCTGA TGATGTTGAA 1020

GCATTTGTTA TTGACGCCGT AAGAACTAAA ATGGTCTACT GCAAAATTGA TCAGACCCAG 1080

AGAAAAGTAG TTGTCAGTCA TAGCACACAT CGGACATTTG GAAAACAGCA GTGGCAACAA 1140

CTGTATGACA CACTTAATGC CTGGAAACAA AATCTGAACA AAGTGAAAAA CAGCCTTTTG 1200

AGTCTTTCTG ATACCTGAGT TTTTATGCTT ATAATTTTTG TTCTTTGAAA AAAAAGCCCT 1260

AAATCATAGT AAAACATTAT AAACTAAAAA AAAAAAAAAA AAAAAAAAAA 1310

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1139 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

```
AGGGCANACT TACAGAGATA TCATATGAGA TCACCCCTCG CATTCGTGTC TGGCGCCAGA      60
CCCTCGAGCG GTGCCGGAGC GCASCCAGGT GTGCTTGTGC CTGGGCCAGC TGGAGAGGTC     120
CATTGCCTGG GANGAAGTCT GTCAACAAAG TGACATGTCT AGTCTGCCGG AAGGGTGACA     180
ATGATGAGTT TCTTCTGCTT TGTGATGGGT GTRACCGTGG CTGCCACATT TACTGCCATC     240
GTCCCAAGAT GGAGGCTGTC CCAGAAGGAG ATTGGTTCTG TACTGTCTGT TTGGCTCAGC     300
AGGTGGAGGG AGAATTCACT CAGAAGCCTG GTTTCCCAAA GCGTGGCCAG AAGCGGAAAA     360
GTGGTTATTC GCTGAACTTC TCAGAGGGTG ATGGCCGCCG ACGCCGGGTA CTGTTGAGGG     420
GCCGAGAAAG CCCAGCAGCA GGGCCTCGGT ACTCGGAAGA AGGGCTCTCC CCCTCCAAGC     480
GGCGGCGACT CTCTATGCGG AACCACCACA GTGATCTCAC ATTTTGCGAG ATTATCCTGA     540
TGGAGATGGA GTCCCATGAT GCAGCCTGGC CTTTCCTAGA GCCTGTGAAC CCACGTTTGG     600
TGAGTGGGTA CCGGCGCATC ATCAAAAATC CTATGGATTT TTCCACCATG CGGGAGCGGC     660
TGCTCAGGGG AGGGTACACC AGCTCAGAGG AGTTTGCGGC TGATGCCCTC CTGGTATTTG     720
ACAACTGCCA GACTTTCAAC GAGGATGACT CTGAAGTAGG CAAGGCTGGG CACATCATGC     780
GCCGCTTCTT CGAGAGCCGC TGGGAGGAGT TTTATCAGGG AAAACAGGCC AATCTGTGAG     840
GCAAGGGAGG TGGGGAGTCA CCTTGTGGCA TCTCCCCCCA CCTTCCAAAC AAAAACCTGC     900
CATTTTCACC TGCTGATGCT GCCCTGGGTC CAGACTCAAG TCAGATACAA CCCTGATTTT     960
TGACCTTNCC CTTGGCAGTG CCCCACATCC TCTTATTCCT ACATCCCTTT CTCCCTTCCC    1020
TCCTCTTGCT CCTCAAGTAA GAGGTGCAGA GATGAGGTCC TTCTGGACTA AAAGCCAAAA    1080
AAAGAAAGAA AAAAAWAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAN     1139
```

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2271 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

GTTCCGGGGG ATGCCAGCTC ACTTCTCGGA CAGCGCCCAG ACTGAGGCCT GCTACCACAT 60
GCTGAGCCGG CCCCAGCCGC CACCCGACCC CCTCCTGCTC CAGCGTCTGC CACGGCCCAG 120
CTCCCTGTCA GACAAGACCC AGCTCCACAG CAGGTGGCTG GACTCGTCGC GGTGTCTCAT 180
GCAGCAGGGC ATCAAGGCCG GGGACGCACT CTGGCTGCGC TTCAAGTACT ACAGCTTCTT 240
CGATTTGGAT CCCAAGACAG ACCCCGTGCG GCTGACACAG CTGTATGAGC AGGCCCGGTG 300
GGACCTGCTG CTGGAGGAGA TTGACTGCAC CGAGGAGGAG ATGATGGTGT TTGCCGCCCT 360
GCAGTACCAC ATCAACAAGC TGTCCCAGAG CGGGGAGGTG GGGGAGCCGG CTGGCACAGA 420
CCCAGGGCTG GACGACCTGG ATGTGGCCCT GAGCAACCTG GAGGTGAAGC TGGAGGGGTC 480
GGCGCCCACA GATGTGCTGG ACAGCCTCAC CACCATCCCA GAGCTCAAGG ACCATCTCCG 540
AATCTTTCGG CCCCGGAAGC TGACCCTGAA GGGCTACCGC CAACACTGGG TGGTGTTCAA 600
GGAGACCACA CTGTCCTACT ACAAGAGCCA GGACGAGGCC CCTGGGGACC CCATTCAGCA 660
GCTCAACCTC AAGGGCTGTG AGGTGGTTCC CGATGTTAAC GTCTCCGGCC AGAAGTTCTG 720
CATTAAACTC CTAGTGCCCT CCCCTGAGGG CATGAGTGAG ATCTACCTGC GGTGCCAGGA 780
TGAGCAGCAG TATGCCCGCT GGATGGCTGG CTGCCGCCTG GCCTCCAAAG GCCGCACCAT 840
GGCCGACAGC AGCTACACCA GCGAGGTGCA GGCCATCCTG GCCTTCCTCA GCCTGCAGCG 900
CACGGGCAGT GGGGGCCCGG GCAACCACCC CCACGGCCCT GATGCCTCTG CCGAGGGCCT 960
CAACCCCTAC GGCCTCGTTG CCCCCCGTTT CCAGCGAAAG TTCAAGGCCA AGCAGCTCAC 1020
CCCACGGATC CTGGAAGCCC ACCAGAATGT GGCCCAGTTG TCGCTGGCAG AGGCCCAGCT 1080
GCGCTTCATC CAGGCCTGGC AGTCCCTGCC CGACTTCGGC ATCTCCTATG TCATGGTCAG 1140
GTTCAAGGGC AGCAGGAAAG ACGAGATCCT GGGCATCGCC AACAACCGAC TGATCCGCAT 1200
CGACTTGGCC GTGGGCGACG TGGTCAAGAC CTGGCGTTTC AGCAACATGC GCCAGTGGAA 1260
TGTCAACTGG GACATCCGGC AGGTGGCCAT CGAGTTTGAT GAACACATCA ATGTGGCCTT 1320
CAGCTGCGTG TCTGCCAGCT GCCGAATTGT ACACGAGTAT ATCGGGGGCT ACATTTTCCT 1380
GTCGACGCGG GAGCGGGCCC GTGGGGAGGA GCTGGATGAA GACCTCTTCC TGCAGCTCAC 1440
CGGGGGCCAT GAGGCCTTCT GAGGGCTGTC TGATTGCCCC TGCCCTGCTC ACCACCCTGT 1500
CACAGCCACT CCCAAGCCCA CACCCACAGG GGCTCACTGC CCCACACCCG CTCCAGGCAG 1560
GCACCCAGCT GGGCATTTCA CCTGCTGTCA CTGACTTTGT GCAGGCCAAG GACCTGGCAG 1620
GGCCAGACGC TGTACCATCA CCCAGGCCAG GGATGGGGGT GGGGGTCCCT GAGCTCATGT 1680

GGTGCCCCCT TTCCTTGTCT GAGTGGCTGA GGCTGATACC CCTGACCTAT CTGCAGTCCC 1740
CCAGCACACA AGGAAGACCA GATGTAGCTA CAGGATGATG AAACATGGTT TCAAACGAGT 1800
TCTTTCTTGT TACTTTTTAA AATTTCTTTT TTATAAATTA ATATTTTATT GTTGGATCCT 1860
CCTCCTTTCT CTGGAGCTGT GCTTGGGGCT ACTCTGACAC TCTGTCTCTT CATCACCAGC 1920
CAAGGAAAGG GGCTTTCCTG ATAAAGACAA GAGTTGGTTA GAGAAAGGGA CACCTAAGTC 1980
AGTCTAGGGT TGGAAGCTAG GAGAGAGGTG AGGGCAGAAG GGCACAGCTT TCAGGAACAA 2040
GGAATAGGGG CTGGGGTKGT KGTTCTCACG GGTAGGCGTA CCTGCAGGGC CTCCTTGAAG 2100
TACTTGGGAA GGAGGAAGCC ATCAGTATTC CCTGGAGTCA GAATCACCCC ATTGGCAGAG 2160
CGGAAGAAGG GTATTCCATC TGCTGACAGA GCCAGAGATG TGACTCATGC CCTCCCCGAA 2220
GGCAAAGTCA GCTCCTGCTT TGTCCAGACT CACCTGCCAG AGCCAGGGGT C 2271

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 626 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

ACAACAAACA TCGAAAATCG ANTATGTGCC CCGAAAAGTC GGAACGCAGG CAATCAGTCC 60
GCACGMGCGC AAGTTCAACA TGAAGATGAT ATGAGGCCGG GGCGGGGGGC AGGGACCCCC 120
GGGCGGCCGG GCAGGGGAAG GGGCCTGGCC GCCACCTGCT CACTCTCCAG TCCTTCCCAC 180
CTCCTCCCTA CCCTTCTACA CACGTTCTCT TTCTCCCTCC CGCCTCCGTC CCCTGCTGCC 240
CCCCGCCAGC CCTCACCACC TGCCCTCCTT CTACCAGGAC CTCAGAAGCC CAGACCTGGG 300
GACCCCACCT ACACAGGGGC ATTGACAGAC TGGAGTTGAA AGCCGACGAA CCGACACGCG 360
GCAGAGTCAA TAATTCAATA AAAAAGTTAC GAACTTTCTC TGTAACTTGG GTTTCAATAA 420
TTATGGATTT TTATGAAAAC TTGAAATAAT AAAAAGAGAA AAAAACTATT TCCTATAGCT 480
AGTCGGAATG CAAACTTTTG ACGTCCTGAT TGCTCCAGGG CCCTCTTTCC AACTCAGTTT 540
CTTGTTTTTC CTCTTCCTCC TCCTCCTCTT CTTCCTCCTT TCTTTCTCTT NCCCCATGGG 600
GGAGGGGTTC ATTCAGGGAA AACAGG 626

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2118 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

TTTTCCAGCC ATGTCACTAA TTGTGAATTC CTACCAACTA TTGACAGAAT ACAGAGTTGA 60
TTTTTTAATA AAAAGTTATA TATAATTATC CCTTTAATTA AAGGGAGCAA AGGGGCGTTC 120
CACATGGACA GAGGCTTGGA CCGAGGCCTG GTCACAGCAG CGAGCATCCA GGGTTTGCAG 180
GGACGATGTT ACAGACTCTG TTTTCTGCCT GGCGTTTCAC TTGTGTCTGC TCCTAGCCTG 240
TGCTCTGCCA GCAGCACAGA CATCTGCTCC ATCAGACCTC TTCCATTTTG CACAGGGAGT 300
GCAGGAGGTG AATGTTCACT TTCTGTTCTC CAGTGTCACT GTTCTGTTTC CACGGGATGG 360
AAAGCGCATG GGCCTGTGTC CATTGTAGAT TTCCTTCTAG ATTTCTGTGT ACACACACTT 420
GATTGTTCTG GATGAATGTC TTTTTTAATA CTCCGAAAAT TTCATCATCT AAGAAAATGA 480
TTCCATACAA ATAACTCAGC ACACAAGTGA CCCAGGACAT ATGCCTGCCA AAGGGATGTG 540
TTAGAAGGCT GCCTTCTCAT GCGCATTGTC ACTTGGATCT TGTGGTGAGG ACGGCCCCAT 600
CTTTCTTGCC ACAGATTGAG GCCACTTTTG AGCAAGGGAG ATCCTGGAGT TAAGACAGGT 660
GTTGGGGGCA GCCTGTATTT TACCCTAGGG GCAGGTCTGC ATGGTGACCC CACATYGCAC 720
TGGTAAACCA TTTGAGTCCC ACTCTTCATC CTGGAAGTGG GAACTGGAGT CCCACCCACA 780
GTGCATTCAG AAAGCATGCT GTGTGGGGGC TGCTTCTCAG GAGGCCAGGC CCTTCTGAGC 840
GGAACCGTCC TGGAGAGAGC CTGCCCTCGT TTCCAGGCTG CAGCCGTAAC GCACTTTCTC 900
CCAGGCTGAG GGCGGGTGTT CTGGGGTGTC TGCCCTCTGT CGGCCCTGCT TCCTGCCAGG 960
ACGTGGCCTC TTCCGATCCT TTTCTCTCAG ACACTGGAGG TCTCTTCTGC CATTGTGCTG 1020
GTCCCATCCC AAGAATTGTA GGACAGAGAC CACACTGGGT CGGCGGACAC AAAGTCCATC 1080
CAGGACCCAG GCCGCAGAGG GAGCAGGAAG AGATGCTGAT AGTTTGATCT AGAAACCAGC 1140
AGCTACTGGC TCAAATTCAG GTTCTGGCGT CAAATAGCGA CATTTCCAGT TTCTCTTAAA 1200
AACCGTGTTT GGTTTCAGTT GGGATAGGCT TGTTTTGTCT GTTGAAAATG TTTCTAGTTT 1260
TTTTTCTTTC ATTTTTCTCT CATTCCATTT CTGCCTTAAC TTTAGTTTGT TCACAGGGAG 1320
GCAAAGCTGA CATGAACCTT TTGTCGTGGG ACTTCAGGCC ACATTGGCTT GAAGGCATTC 1380
GTTTCCTTCT GGGGTGGGGA CAGGCCCTCA TGGCAGGCTT GTTCCCGTGG CTCTGAGCGA 1440
GGCCTCTTCC TGCTGGGCTC CCAGACTCCT GCATCCAGGC CCCCACCTTC TCGGCTTCTG 1500
GTTTTTCTTT CTTTTTGGTA GAACACAACA TCTACCATTC AGTTAAACCT TCTTTATCTC 1560

CTCCTYTGGC ATCCATTTTT CCAAAGAAGA GTCGAGTCCT CTGAGGTCTG TGCTTGAAAR 1620

CCGTCCGAAG GCATTCTTGT TAGCTTTGCT TTTCTCCCCA TATCCCAAGG CGAAGCGCTG 1680

AGATTCTTCC ATCTAAAAAA CCCTCGACCC GAAACCCTCA CCAGATAAAC TACAGTTTGT 1740

TTAGGAGGCC CTGACCTTCA TGGTGTCTTT GAAGCCCAAC CACTCGGTTT CCTTCGGATT 1800

TTCCTCCCTT TGTTCGGGGT TTGGTTTGGC TCCTCTGTGT GTGTCCGTAT CTTGTTCGGT 1860

GTCCTCGAGG TTGAGCTTCA CTCCACTGCG GCAGAGGCAG CGTGCACACT CGGATTTGCT 1920

ACGTTTCTAT ATATCTTGAA GCTAAATGTA TATATGAGTA GTTTGCCATG AGATAACACA 1980

GTGTAAACAG TAGACACCCA GAAATCGTGA CTTCTGTGTT CTCTCCATTT GAGTATTTTG 2040

TAATTTTTTT GAAATATTTG TGGACATAAA TAAAACCAAG CTACACTACA AAAAAAAAAA 2100

AAAAAAACTG GAGACTAG 2118

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1076 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

GCCCAAGGAG CTCAGCTTCG CCCGCATCAA GGCCGTTGAG TGCGTGGAGA GCACCGGGCG 60

CCACATCTAC TTCACGCTGG TGACCGAAGG GWGCGGCGAG ATCGACTTCC GCTGCCCCCT 120

GGAAGATCCC GGCTGGAACG CCCAGATCAC CCTAGGCCTG GTCAAGTTCA AGAACCAGCA 180

GGCCATCCAG ACAGTGCGGG CCCGGCAGAG CCTCGGGACC GGGACCCTCG TGTCCTAAAC 240

CACCGGGCGC ACCATCTTTC CTTCATGCTA CCCACCACCT CAGTGCTGAG GTCAAGGCAG 300

CTTCGTTGTT CCCTCTGGCT TGTGGGGGCA CGGCTGTSYT CCATGTGGCA AGGTGGAAGG 360

CATGGACGTG TGGAGGAGGC GCTGGAGCTG AAGGAATGGA CGAGCCCTGG GAGGAGGGCA 420

GAAGGCTACG CAGGGCTGAG GATGAAGATG CAGCCCCTGG ATGGTCCCAG ACTCTCAGGA 480

CATGCCCAGC TCAGGGGCTT CGAGCCACAG GCCTGGCCTC ATATGGCATG AGGGGGAGCT 540

GGCATAGGAG CCCCCTCCCT GCTGTGGTCC TGCCCTCTGT CCTGCAGACT GCTCTTAGCC 600

CCCTGGCTTT GTGCCAGGCC TGGAGGAGGG CAGTCCCCCA TGGGGTGCCG AGCCAACGCC 660

TCAGGAATCA GGAGGCCAGC CTGGTACCAA AAGGAGTACC CAGGGCCTGG TACCCAGGCC 720

CACTCCAGAA TGGCCTCTGG ACTCACCTTG AGAAGGGGGA GCTGCTGGGC CTAAAGCCCA 780

CTCCTGGGGG TCTCCTGCTG CTTAGGTCCT TTTGGGACCC CCACCCATCC AGGCCCTTTC 840

```
TTTGCACACT TCTTCCCCCA CCTCTAYGCA TCTTCCCCCC ACTGCGGTGT TCGGCCTGAA      900
GGTGGTGGGG GTGAGGGGGG GTTTGGCCAT TAGCATTTCA TGTCTTTCCC CAAATGAAGA      960
TGCCCTGCAA AGGGCAGTNA ACCACAAAAA AAAAAAAAAA AAAAACNTGG GGGGGGGGCC     1020
CCGTTAACCA TTTTGGCCTN ATAGGGGGGN GGTTTTTAAA AATTAATTGG GCCCGG         1076
```

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1379 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

```
GGCACGAGCA CCCTCCCACA CCTCCCTGAA CTTCCATCTG ATCGACTTCA ACTTGCTGAT       60
GGTGACCACC ATCGTTCTGG GCCGCCGCTT CATTGGGTCC ATCGTGAAGG AGGCCTCTCA      120
GAGGGGGAAG GTCTCCCTCT TTCGCTCCAT CCTGCTGTTC CTCACTCGCT TCACCGTTCT      180
CACGGCAACA GGCTGGAGTC TGTGCCGATC CCTCATCCAC CTCTTCAGGA CCTACTCCTT      240
CCTGAACCTC CTGTTCCTCT GCTATCCGTT TGGGATGTAC ATTCCGTTCC TGCARCTGAA      300
TTKCGAMCTY CGSAAGACAA GCCTCTTCAA CCACATGGCC TCCATGGGGC CCCGGGAGGC      360
GGTCAGTGGC CTGGCAAAGA GCCGGGACTA CCTCCTGACA CTGCGGGAGA CGTGGAAGCA      420
GCACASAAGA CAGCTGTATG GCCCGGACGC CATGCCCACC CATGCCTGCT GCCTGTCGCC      480
CAGCCTCATC CGCAGTGAGG TGGAGTTCCT CAAGATGGAC TTCAACTGGC GCATGAAGGA      540
AGTGCTCGTS AGCTCCATGC TGAGCGCCTA CTATGTGGCC TTTGTGCCTG TYTGGTTCGT      600
GAAGAACACA CATTACTATG ACAAGCGCTG GTCCTGTGNA ACTCTTCCTG CTGGTGTCCA      660
TCAGCACCTC CGTGATCCTC ATGCAGCACC TGCTGCNTGC CAGCTACTGT GACCTGCTGC      720
ACAAGGCCGC CGCCCATCTG GGCTGTTGGC AGAAGGTGGA CCCAGCGCTG TGCTCCAACG      780
TGCTGCAGCA CCCGTGGACT GAAGAATGCA TGTGGCCGCA GGGCGTGCTG GTGAAGCACA      840
GCAAGAACGT CTACAAAGCC GTAGGCCAMW ACAAMGTGGC TATCCCCTCT GACGTCTCCC      900
ACTTCCGCTT CCAKTTCTTT TTCAGCAAAC CCCTGCGGAT CCTCAACATC CTCCTGCTGC      960
TGGAGGGCGC TGTCATTGTC TATCAGCTGT ACTCCCTAAT GTCCTCTGAA AAGTGGCACC     1020
AGACCATCTC GCTGGCCCTC ATCCTCTTCA GCAACTACTA TGCCTTCTTC AAGCTGCTCC     1080
GGGACCGCTT GGTATTGGGC AAGGCCTACT CATACTCTGC TAGCCCCCAG AGAGACCTGG     1140
```

ACCACCGTTT CTCCTGAGCC CTGGGGTCAC CTCAGGGACA GCGTCCAGGC TTCAGCAAGG 1200

GCTCCCTGGC AAGGGGCTGT TGGGTAGAAG TGGTGGTGGG GGGGACAAAA GACAAAAAAA 1260

TCCACCAGAG CTTTGTATTT TTGTTACGTA CTGTTTCTTT GATAATTGAT GTGATAAGGA 1320

AAAAAGTCCT ATTTTTATAC TCCCAANMAA AAAAAAAAAA NAAAAAGCGG CCGAAAGCT 1379

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1337 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

CTGGTGTTGG GCCTGAGCCN CCTCAACAAC TCCTACAACT TCAGTTTCCA CGTGGTGATC 60

GGCTCTCAGG CGGAAGAAGG CCAGTACAGC CTGAACTTCC ACAACTGCAA CAATTCAGTG 120

CCAGGAAAGG AGCATCCATT CGACATCACG GTGATGATCC GGGAGAAGAA CCCCGATGGC 180

TTCCTGTCGG CAGCGGAGAT GCCCCTTTTC AAGCTCTACA TGGTCATGTC CGCCTGCTTC 240

CTGGCCGCTG GCATCTTCTG GGTGTCCATC CTCTGCAGGA ACACGTACAG CGTCTTCAAG 300

ATCCACTGGC TCATGGCGGC CTTGGCCTTC ACCAAGAGCA TCTCTCTCCT CTTCCACAGC 360

ATCAACTACT ACTTCATCAA CAGCCAGGGG CCACCCCATC GAAGGCCTTG CCGKCATGTA 420

CTACATCGCA CACCTGCTGA AGGGCGCCCT CCTCTTCATC ACCATCGCCC TGATTGGCTC 480

AGGCTGGGCT TCATCAAGTA CGTCCTGTCG GATAAGGAGA AGAAGGTCTT TGGGATCGTG 540

ATCCCCATGC AGGTCCTGGC CAACGTGGCC TACATCATCA TCGAGTCCCG CGAGGAAGGC 600

GCCACGAACT ACGTGCTGTG GAAGGAGATT TTGTTCCTGG TGGACCTCAT CTGCTGTGGT 660

GCCATCCTGT TCCCCGTAGT CTGGTCCATC CGGCATCTCC AGGATGCGTC TGGCACAGAC 720

GGGAAGGTGG CAGTGAACCT GGCCAAGCTG AAGCTGTTCC GGCATTACTA TGTCATGGTC 780

ATCTGCTACG TCTACTTCAC CCGCATCATC GCCATCCTGC TGCAGGTGGC TGTGCCCTTT 840

CAGTGGCAGT GGCTGTACMA GCTCTTGGTG GARGGCTCCA CCCTGGCCTT CTTCGTGCTC 900

ACGGGCTACA AGTTCCAGCC CACAGGGAAC AACCCGTACC TGCAGCTGCC CCAGGAGGAC 960

GAGGAGGATG TTCAGATGGA GCAAGTAATG ACGGACTCTG GGTTCCGGGA AGGCCTCTCC 1020

AAAGTCAACA AAACAGCCAG CGGGCGGGAA CTGTTATGAT CACCTCCACA TCTCAGACCA 1080

AAGGGTCGTC CTCCCCCAGC ATTTCTCACT CCTGCCCTTC TTCCACAGCG TATGTGGGGA 1140

GGTGGAGGGG TCCATGTGGA CCAGGCGCCC AGCTCCCGGG ACSCCGGTTC CCGGACAAGC 1200

CCATTTGGAA GAAGAGTCCC TTCCTCCCCC CAAATATTGG GCAGCCCTGT CCTTACCCCG 1260

GGACCACCCC TCCCTTCCAG CTATGTGTAC AATAATGACC AATCTGTTTG GCTAAAAAAA 1320

AAAAAAAAAA AACTCGA 1337

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1390 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

GCCGTTTTGG TTCCCGGTTG GTGCTTCCTG TTCGCAGCTG CGGCACTTCA AGGTTACTGA 60

CTTTTTATGA TGTTTGGTGG CTATGAGACT ATAGAWGCRT RSGRRGATGA TYTTTATCGA 120

GATGAGTCAT CTAGTGAACT GAGTGTTGAT AGTGAGGTGG AATTTCAACT CTATAGCCAA 180

ATTCATTATG CCCAAGATCT TGATGATGTC ATCAGGGAGG AAGAGCATGA AGAAAAGAAC 240

TCTGGGAATT CGGAATCTTC GAGTAGTAAA CCAAATCAGA AGAAGCTAAT CGTCCTTTCA 300

GATAGTGAGG TCATCCAGCT GTCAGATGGG TCAGAGGTCA TCACTTTGTC TGATGAAGAC 360

AGTATTTATA GATGTAAAGG AAAGAATGTT AGAGTTCAAG CACAAGAAAA TGCCCATGGT 420

CTTTCTTCTT CTCTTCAATC TAATGAGCTG GTTGATAAGA AATGCAAGAG TGATATTGAG 480

AAGCCTAAAT CTGAAGAGAG ATCAGGTGTA ATCCGAGAGG TCATGATTAT AGAGGTCAGT 540

TCAAGTGAAG AGGAAGAGAG CACCATTTCA GAAGGTGATA ATGTGGAAAG CTGGATGCTA 600

CTGGGATGTG AAGTAGATGA TAAAGATGAT GATATCCTTC TCAACCTTGT GGGATGTGAA 660

AACTCTGTTA CTGAAGGAGA AGATGGTATA AACTGGTCCA TCAGTGACAA AGACATTGAG 720

GCCCAGATAG CTAATAACCG AACACCTGGA AGATGGACCC AGCGGTACTA TTCAGCCAAC 780

AAAAACATTA TCTGTAGAAA TTGTGACAAA CGTGGTCATT TATCAAAAAA CTGCCCCTTA 840

CCACGAAAAG TTCGTCGCTG CTTCCTGTGC TCCAGGAGAG GACATCTCCT GTATTCCTGT 900

CCAGCCCCCC TTTGCGAATA CTGTCCTGTG CCTAAGATGT TGGACCACTC ATGTCTTTTC 960

AGACATTCCT GGGATAAACA GTGTGACCGA TGTCATATGC TAGGCCACTA TACAGATGCT 1020

TGCACAGAAA TCTGGAGGCA GTATCACCTA ACGACCAAAC CTGGACCACC CAAAAAGCCG 1080

AAGACCCCTT CAAGACCATC AGCCTTAGCA TATTGCTATC ACTGCGCGCA AAAAGGCCAT 1140

TATGGACACG AATGTCCAGA AAGAGAAGTG TATGACCCGT CTCCAGTATC TCCATTCATC 1200

```
TGCTACTATG RTGACAAATA TGAAATTCAG GAGAGAGAAA AGAGACTAAA ACAAAAAATA   1260
AAAGTANTCA AGAAAAATGG GGTTATCCCA GAGCCATCCA AGCTACCTTA TATAAAAGCA   1320
GCAAATGAGA ACCCCCACCA TGATATAAGG AAGGGCCGTG CCTCATGGAA AAGCAACAGG   1380
TGGCCTCAAG                                                          1390
```

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1431 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

```
GCCTGCAGTC GACACTAGTG GATCCAAAGA ATTCGGCCTG TGCGAGTAGG CGCTTGGGCA     60
CTCAGTCTCC CTGGCGAGCG ACGGGCAGAA ATCTCGAACC AGTGGAGCGC ACTCGTAACC    120
TGGATCCCAG AAGGTCGCGA AGGCAGTACC GTTTCCTCAG CGGCGGACTG CTGCAGTAAG    180
AATGTCTTTT CCACCTCATT TGAATCGCCC TCCCATGGGA ATCCCAGCAC TCCCACCAGG    240
GATCCCACCC CCGCAGTTTC CAGGATTTCC TCCACCTGTA CCTCCAGGGA CCCCAATGAT    300
TCCTGTACCA ATGAGCATTA TGGCTCCTGC TCCAACTGTC TTAGTACCCA CTGTGTCTAT    360
GGTTGGAAAG CATTTGGGCG CAAGAAAGGA TCATCCAGGC TTAAAGGCTA AAGAAAATGA    420
TGAAAATTGT GGTCCTACTA CCACTGTTTT TGTTGGCAAC ATTTCCGAGA AAGCTTCAGA    480
CATGCTTATA AGACAACTCT TAGCTAAATG TGGTTTGGTT TTGAGCTGGA AGAGAGTACA    540
AGGTGCTTCC GGAAAGCTTC AAGCCTTCGG ATTCTGTGAG TACAAGGAGC CAGAATCTAC    600
CCTCCGTGCA CTCAGATTAT TACATGACCT GCAAATTGGA GAGAAAAAGC TACTCGTTAA    660
AGTTGATGCA AAGACAAAGG CACAGCTGGA TGAATGGAAA GCAAAGAAGA AAGCTTCTAA    720
TGGGAATGCA AGGCCAGAAA CTGTCACTAA TGACGATGAA GAAGCCTTGG ATGAAGAAAC    780
AAAGAGGAGA GATCAGATGA TTAAAGGGGC TATTGAAGTT TTAATTCGTG AATACTCCAG    840
TGAGCTAAAT GCCCCCTCAC AGGAATCTGA TTCTCACCCC AGGAAGAAGA AGAAGGAAAA    900
GAAGGAGGAC ATTTTCCGCA GATTTCCAGT GGCCCCACTG ATCCCTTATC CACTCATCAC    960
TAAGGAGGAT ATAAATGCTA TAGAAATGGA AGAAGACAAA AGAGACCTGA TATCTCGAGA   1020
GATCAGCAAA TTCAGAGACA CACATAAGAA ACTGGAAGAA GAGAAAGGCA AAAAGGAAAA   1080
AGAAAGACAG GAAATTGAGA AAGAACGGAG AGAAAGAGAG AGGGAGCGTG AAAGGGAACG   1140
AGAAAGGCGA GAACGGGAAC GAGAAAGGGA AAGAGAACGT GAACGAGAAA AGGAGAAAGA   1200
```

```
ACGGGAGCGG GAACGAGAAC GGGATAGGGA CCGTGACCGG ACAAAAGAGA GAGACCGAGA   1260
TCGGGATCGA GAGAGAGATC GTGACCGGGA TAGAGAAAGG AGCTCAGATC GTAATAAGGA   1320
TCGCATTCGA TCAAGAGAAA AAAGCAGAGA TCGTGAAAGG GAACGAGAGC GGGAAAGAGA   1380
GAGAGAGAGA GAACGAGAGC GAGAACGAGA ACGGGAGCGA GAGAGAGAAG C            1431
```

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2539 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

```
GGGTGCAGGA GTGCCACCCC CAGGGCCCTG TCAACCTCTC TTTTCTCCTC CATGGCTGTC     60
TGCCTGCGTA TCTGTCTCTG AGAATCCTCG GGGCGGTCAG GGGATGTCAG GAGGGGAAGG    120
AGCCGCCCTC CCTATCTTGC TGCTCCTCTT GGCACTCAGG GGCACCTTCC ATGGAGCCAG    180
ACCGGGTGGA GGGGCTTCTG GGATTTGGTG TCTGCTGCTG CCAGAGCAGG AACCCCCAGT    240
CTAGGACTTG GGCATTTTAA CAGGGAGAAA GTAGTGGCTT CCCTTTTCTC TCTCTCCTCC    300
TTTTTCCCTT TAAGCCCACA GATTCAGGTC ATGCCAAAAG CTCTCTGGTT GTAACCTGGA    360
GACATGTGGA GGGGAATGGC GATGGGATTA TAGGACTCTC CCCATCTCGG GCCCTGACCC    420
TGACCCTTGC CACCAACCCA AAGACAGCTG GTGGGTTTCC CCTTGGAGAM AATCCTGCGT    480
TTGCCTGGGC CGGCCCTGGC TGCCCTCAGC TTTCGCTGAT CTGCCCGGCC TGGAGCCTCC    540
CATCACCCCG CTTCTTGTTG GGCCTCAGGC ACTGGTTACC AGAAGGGGGT CTGGGTCTGC    600
TCAGGAATCA TGTTTTGTAG CACCTCCTGT TGGAGGGGTG GAGGGATGTT CCCCTGAGCC    660
AGGCTGAGAC TAGAACCCCA TCTTCCCTGA GCCAGGCTGA GACTAGAACC CCATCTTCCC    720
CACCACGCCA CCCCTGTGST KGCTACAGGA GCACAGTAGT GAAGGCCTGA GCTCCAGGTT    780
TGAAAGACCC AACTGGAGCG TGGGGCGGGC AGGCAGGGGT TAGTGAAAGG ACACTTCCAG    840
GGTTAGGACA GAGCATTTAG CCTTCTGGAA GAACCCCTGC CTGGGGTGGG ACTGTGCAGG    900
CCAGAGAAGG TGGCATGGGC CTGAACCCAC CTGGACTGAC TTCTGCACTG AAGCCACAGA    960
TGGAGGGTAG GCTGGTGGGT GGGGGTGGTT CGTTCTCTAG CCGGGGCAGA CACCCAGCTG   1020
GCTGGGTCCT TCCTCAGCCT TGCCTCCTCC TGTCCCCAAC CCTTTCCTTT CCTCCTGCTT   1080
GCGGACTGCT GGTCCCCTCT CCTTCCCTCC TTCCAGCTGT TTCTAGTTAC CACCTACCCC   1140
```

```
TGGGCCGTGG ACTGATCAGA CCAGCATTCA AAATAAAAGT TTGTTCCAAG TTGACAGTGT    1200
GGTGCTCCCT GCCCAGCCCC TCCAGGTGGA GGTGCTGCCA CGGGAACGCA GTTGCTCTGC    1260
CTGCCCTGGG CCCCTGGCGA CANTGGGAGC AGGGCAGTGC TGTGAGGAGC CCAGCTTTCC    1320
CAGTCAGGCA GGCATGGCTT CCGTGTTCAG GCTCCCTCAC CAGCTGGTGA CACGGGACAA    1380
GCTTACAAAC CTTCTCTGAA CCTCAGTTTT CTCATTTACA AGAGGCAAAG CATCCATCAC    1440
CTTGTGTGGA TTCARAGAAT GTRAGGCCCT GGGGTGTCCT ACACAAGGGA AAGGCTTGCT    1500
CAGTGAGCGG TCTGCACACC GTTAGCCACC CTGCCACCTC TGTGCCCTGG GCAGGCTCCA    1560
AAGGAAAGCT CTGGCTGGGA CTGCCRGGAG TCTCACACGC TCCTGTTGAC ATTCCCAGCA    1620
GCYGCCCCTG AGGTCGATGT TTGTTCTGTT TTTCTTTTTC TTTTTTGAGA CGGAGTCTCG    1680
CTGTGTTGCC AGGCTGGAGT GCAGTGGTGT GATCTCTGCT CACTGCAACC TCCGCCTGCC    1740
AGTTTCAAGT GATTCTCTGC CTCAGCCTTC TGAGTAGCTG GGACTACAGG TGCACGCCAC    1800
CACGCCCAGC TAACTTTTTG TATTTWAGTA GAGACAGGGT TTCGCCATGT CGGCCAGGGT    1860
GGTCTTGATC TCCTGACCTC ATGATCCACC CGCCTCAGCC TCCCAAAGTG CTGGGATTAC    1920
AGGTATGAGC CACCGCACCG GGCCTGTTCT ATTTTTCTAG TTAAGGGAAC TGAAGCTCAG    1980
ARAGGTGTCA CCAGCARGTG TTCATTCCCA TGCCAGCCTT GCCCCCCGGC TTTTCCCAGG    2040
CAGGCTCCTG CGTGCCCACT GGCTCCAGCC TGGTCCTCTG TCTCTTGGCT GCTTCACTCC    2100
TGCTCTTTGT CCCGACTCTG GCCCTGCTTA CAGGGGCCAC TACCTGCTGG TGCCTCCATA    2160
ACAAGCGTCT GGCGTTGAGA CCCCTGGCAT GGCAGGGGCT TTGGGGTCTG GTTTCCACAA    2220
GGCTTAGCCA TGGCAGAACC TCGTTTTATT TTAACTCTTT GCCCCTACAA ACAAACAGCA    2280
GTACTTGCCA GAACCATTCT TGGGATTCAG GAGCTCGGGC GACTGCCTTG GCCTCTGGCC    2340
GCACCCAGGA GGGTGGGGTT GGATCTGTGT AGTTGCCAGG CCCACACCTG CCAGCAGGGG    2400
GCTGACTGGA TCCATGCTTT ACTGTGTTTA ATGGGGGTAA CAGGGGTCCC TACAGCCCTC    2460
CCAGYTAAAM ATTTGGAACA AAACACCAGC CCTTTTGTAG TGGATGCAGA ATAAAATTGT    2520
TAATCCAATC AAAAAAAAA                                                 2539
```

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1041 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

TCGACCCACG CGTCCGCCCA CGCGTCCGCC CACGCGTCCG GGCGCAGGAC GTGCACTATG 60

GCTCGGGGCT CGCTGCGCCG GTTGCTGCGG CTCCTCGTGC TGGGGCTCTG GCTGGCGTTG 120

CTGCGCTCCG TGGCCGGGGA GCAAGCGCCA GGCACCGCCC CCTGCTCCCG CGGCAGCTCC 180

TGGAGCGCGG ACCTGGACAA GTGCATGGAC TGCGCGTCTT GCAGGGCGCG ACCGCACAGC 240

GACTTCTGCC TGGGCTGCGC TGCAGCACCT CCTGCCCCCT TCCGGCTGCT TTGGCCCATC 300

CTTGGGGGCG CTCTGAGCCT GACCTTCGTG CTGGGGCTGC TTTCTGGCTT TTTGGTCTGG 360

AGACGATGCC GCAGAGAGAG AAGTTCACCA CCCCCATAGA GGAGACCGGC GGAGAGGGCT 420

GCCCAGCTGT GGCGCTGATC CAGTGACAAT GTGCCCCCTG CCAGCCGGGG CTCGCCCACT 480

CATCATTCAT TCATCCATTC TAGAGCCAGT CTCTGCCTCC CAGACGCGGC GGGAGCAAGC 540

TCCTCCAACC ACAAGGGGGG TGGGGGGCGG TGAATCACCT CYGAGGCCTG GGCCCAGGGT 600

TCAGGGGAAC TTCCAAGGTG TCTGGTTGCC CTGCCTCTGG CTCCAGAACA GAAAGGGAGC 660

CTCACGCTGG CTCACACAAA ACAGCTGACA CTGACTAAGG AACTGCAGCA TTTGCACAGG 720

GGAGGGGGGT GCCCTCCTTC CTAGAGGCCC TGGGGGCCAG GCTGACTTGG GGGGCAGACT 780

TGACACTAGG CCCCACTCAC TCAGATGTCC TGAAATTCCA CCACGGGGGT CACCCTGGGG 840

GGTTAGGGAC CTATTTTTAA CACTAGGGGG CTGGCCCACT AGGAGGGCTG GCCCTAAGAT 900

ACAGACCCCC CCAACTCCCC AAAGCGGGGA GGAGATATTT ATTTTGGGGA GAGTTTGGAG 960

GGGAGGGAGA ATTTATTAAT AAAAGAATCT TTAACTTTAA AAAAAAAAAA AAAAAAGGGC 1020

GGCCGCTCTA GAGGATCCCT C 1041

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1962 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

ACCCACGCGT CCGGTACAAA ACACAGTTTT ATTCTATGAA AATTTTGAGA TTATTAGAAA 60

CATTAGATTT AGGGTTGCAT ATTAAAAACT ATATCCATTT TGCCTTATTA TTTAGTGTCT 120

CACTCAGGAT ATAACACACT ATAATAGAAA ATGTAGACTT CAGAATCAGG TATATTTGAG 180

ATGGTTTGTA TACTGGTTCT GACACTTGTT AGCTATTCAT CTTTGGTAAA TTCCCCATTA 240

CCCTTTGTKC ACCTATWTGT GGGGATCAGT GCATAGTGTG TGTWAAGCAT TTAATACCTG 300

GCAAGTGTTC AGCAAATTTT TTGTTCTATA TATTTATTAT TTGATTATTG GCCCTGAGGA 360
GTAGGTGTTT GTTTGTTTGT TTGTTTGTTT AGTTTTATTT CTCATCTCCT CAGGAACACA 420
AATGAAACTT GGATATTGTT ATGGTGCTTT TNATAATATA TTTATTATTT TCAGCAATTN 480
ATTCTTGTTA AAACAATTTC TTATGACAAG TTACTCATCT TCAATGGTGA GAAGAAATCT 540
AGCTCAGAAT AATATATTTT TAGTGTTTGT ATCTCTGGAT ACTCATTTTG CTCATTGCCA 600
CGTAAAGTAA AAAAATACAT AAATTAGCTT ATTCCAATGT AATATCTTCA GGATAGTCAT 660
GGGCAAGGAA TTAATCACAT TAAGAGATAA CTGCAACTAA GCACTATTTG AGGTGACTTC 720
TGTGGAAAAA AAATTAATYC TTTACCATTG CAGCGTTCTG CCCTAGGTCC AAATGTTACC 780
AAAATCACTC TAGAATCTTT TCTTGCCTGG AAGAAAAGGA AAAGACAAGA AAAGATTGAT 840
AAACTTGAAC AAGATATGGA AAGAAGGAAA GCTGACTTCA AAGCAGGGAA AGCACTAGTG 900
ATCAGTGGTC GTGAAGTGTT TGAATTTCGT CCTGAACTGG TCAATGATGA TGATGAGGAA 960
GCAGATGATA CCCGCTACAC CCAGGGAACA GGTGGTGATG AGGTTGATGA TTCAGTGAGT 1020
GTAAATGACA TAGATTTAAG CCTGTACATC CCAAGAGATG TAGATGAAAC AGGTATTACT 1080
GTAGCCAGTC TTGAAAGATT CAGCACATAT ACTTCAGATA AAGATGAAAA CAAATTAAGT 1140
GAAGCTTCTG GAGGTAGGGC TGAAAATGGT GAAAGAAGTG ACTTGGAAGA GGACAACGAG 1200
AGGGAGGGAA CGGAAAATGG AGCCATTGAT GCTGTTCCTG TTGATGAAAA TCTTTTCACT 1260
GGAGAGGATT TGGATGAACT AGAAGAAGAA TTAAATACAC TTGATTTAGA AGAATGACAC 1320
CAAACACATC GCTGAAAAAA TTAAGTCAGC TCAGCACGAG TTGAAATTGA CTACATTAAT 1380
TTCTTTCCAC CTAGAATCAA CAGGATGTTT ATTTCCTATG CTGATTCTGG AGGAGTTAAC 1440
CTCCTGCAAA AAAGGCATCT TGTCCCTACA TCTTCTCTTC TGACTTTGGC TACATCTCAT 1500
AGTAAGTTCA GAGTAGTTCA TGATAAATTG AAAATATAAT GGTCATTGCA GAAAATGATT 1560
GATGTTGTAA CTGTCCACCC AAGTAAGAAG TGTATCTGCC TTTCCATCTT TTGGTTTTCA 1620
TTTGGGCATG TGCTATTACC AGAAACAACA AACTTATATT TAAAATACCC TTCATTTGAC 1680
ACAGTTTTTA ATGAGTGATT TAATTTCCTC TGTATTTGTA TGTTTAGAAG ACTGCCTAAA 1740
ACATGAGCAC TGTACTTCAT AAAGGAAACG CGTATGCAGA TTCAGTATTG TGTATCTTTG 1800
GACAATTAGA TGGACATTTA AAATGGAACT TCTTTTATCT GACAGGATCA GCTACAATGC 1860
CCTGTGTTAA ATTGTTTAAA AGTTTCCCTT TTCTTTTTTG CCAATAAAGT TGTAAATAAA 1920
GACCATCATA CATTAAAATC CAAAAAAAAA AAAAAAAAAA AA 1962

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1228 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

```
GGCTGCCCAG GCCCCGCACT GGAAGAGCCT CCAGCAGCAA GATGTGACCG YTGTGCCGAT     60
GAGCCCCAGC AGCCACTCCC CAGAGGGGAG GCCTCCACCT CTGCTGCCTG GGGGTCCAGT    120
GTGTAAGGCA GCTGCATCTG CACCGAGCTC CCTCCTGGAC CAGCCGTGCC TCTGCCCCGC    180
ACCCTCTGTC CGCACCGCTG TTGCCCTGAC AACGCCGGAT ATCACATTGG TTCTGCCCCC    240
TGACATCATC CAACAGGAAG CGTCACCCTG AGGGAGGAGA CAGAAGCCTG GGCCAGGTGA    300
ACAGTGGTAT AGCAGCCACT CCAGCCTCTG CTGCAGCAGC CACCCTGGAT GTGGCTGTTC    360
GGAGAGGCCT GTCCCACGGA GCCCAGAGGC TGCTGTGCGT GGCCCTGGGA CAGCTGGACC    420
GGCCTCCAGA CCTCGCCCAT GACGGGAGGA GTCTGTGGCT GAACATCAGG GGCAAGGAGG    480
CGGCTGCCCT ATCCATGTTC CATGTCTCCA CGCCACTGCC AGTGATGACC GGTGGTTTCC    540
TGAGCTGCAT CTTGGGCTTG GTGCTGCCCC TGGCCTATGN TTCCAGCCTG ACCTGGTGCT    600
GGTGGCGCTG GGGCCTGCCA NTGCCTGCAG GGCCCCCACG CTGCACTCCT GGCTGCAATG    660
CTTCGGGGGC TGGCAGGGGG CCGAGTCCTG GCCCTCCTGG AGGAGAACTC CACACCCCAG    720
CTAGCAGGGA TCCTGGCCCG GGTGCTGAAT GGAGAGGCAC CTCCTAGCCT AGGCCCTTCC    780
TCTGTGGCCT CCCCAGAGGA CGTCCAGGCC CTGATGTACC TGAGAGGGCA GCTGGAGCCT    840
CAGTGGAAGA TGTTGCAGTG CCATCCTCAC CTGGTGGCTT GAAATCGGCC AAGGTGGGAG    900
CATTTACACC GCAGAAATGA CACCGCACGC CAGCGCCCCG CGGCCGCGAT CCGGACCCCA    960
AGCCCACGGC TCCCTCGACT CTGGGGCACG GAACCCCGCC CACTCCCAAT CCCCGCGCCC   1020
CGCCCTCTCC CACCCGTGCT TCCCCCGCTC CACCCCTCAC CTCACCTCGC CCCSGCCCCA   1080
CCCATCGCGC CCCGGCCCGT CCCATCGAGG CCCATGCAAC CCACGCTCGG TYCCGTTCCG   1140
GCCCCTGCGC TCKCGCTKNS TTCGCTCCCC GCCCTTGCGC CGTTAGTAAA CATCGCTCAA   1200
ACGAAAAAAA AAAAAAAAAA AAACTCGA                                      1228
```

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1340 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

```
AATTCGGCAG AGAGATGGCC GCCCCCGTGG ATCTAGAGCT GAAGAAGGCC TTCACAGAGC     60
TTCAAGCCAA AGTTATTGAC ACTCAACAGA AGGTGAAGCT CGCAGACATA CAGATTGAAC    120
AGCTAAACAG AACGAAAAAG CATGCACATC TTACAGATAC AGAGATCATG ACTTTGGTAG    180
ATGAGACTAA CATGTATGAA GGTGTAGGAA GAATGTTTAT TCTTCAGTCC AAGGAAGCAA    240
TTCACAGTCA GCTGTTAGAG AAGCAGAAAA TAGCAGAAGA AAAAATTAAA GAACTAGAAC    300
AGAAAAAGTC CTACCTGGAG CGACGTTAAA GGAAGCTGAG GACAACATCC GGGAGATGCT    360
GATGGCACGA AGGGCCCAGT AGGGAGCCTC TCTGGGAAGC TCTTCCTCCT GCCCCTCCCA    420
TTCCTGGTGG GGGCAGAGGA GTGTCTGCAG GGAAACAGCT TCTCCTCTGC CCCGATGGAT    480
GCTTTATTTG GATGGCCTGG CAACATCACA TTTTCTGCAT CACCCTGAGC CCCATTTGCT    540
TCCCAGCCCT GGAGTTTTTA CCCGGCTTTG CTGCCACCTC TGCCCAGGAC ACKCTTCCCT    600
CTCGGGATGT GTGATGAACT CCCAGGAGAG GGAAGATGGG AGCCAGGGCA AGATAGGAAG    660
CTCTGCCTGA GCTTTCCACT AGGCACGCCA GCCAGACCAA TAAAAAGCGT CTGTCCCACT    720
CTGCTAAGCC TGGTTTTCTT GAGCAGAGGG ATGGAACAGA GGGTGAGAGA GGCAGTGGCC    780
GTCTCCACCT CAGCTCCTGC TCCCTCTGCA TCAGAGCCCT TCCTTTCTTG GGGGATGGGC    840
CTTGCCNTCT TCTCTTTTCC CTTCCTGTAC CTTTGACTAA CGCTCAGCTT CCGGGCCTGC    900
ATGCAGTAGA CAGAAGAGGA AGAAAGAACA GATGTTCACA GCTGAATCTC AGTGAACAGA    960
ATAGCAGTCC CTGGATGGCA GTCTGCCTAA AGATTCCTTT CCCTGCCTTC TCCCATACAT   1020
TCCAAAAGGA AGTTCAACAG TAAGCAGCAC CTCCAAGACT GTCTCCTTTY GGCCARTATC   1080
ATAAGATGGA CGCCATAATC CTGAGGCCTC CTAGAGGCTG AGGGGGCAAC GGTGTGATCC   1140
AGCTGGCTCA TCCCAGCCAG GTGGGCCAAT TATTCAATTT TCAAGAATTT TGTTGCAAGC   1200
CAGTTGTCAA ACACAGCCAT TATAATTATG TAAATTTGCA AATTATGTTA AAAACAAGGA   1260
CAATAAATAT TCAAAATGCA TCCCTAAWWA AAAAAAAAAA AANGGGNGGC CGCNCTAGGG   1320
GATCCAAGCT TACGTACGCG                                               1340
```

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 806 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

```
ACCTTCTTCC ATGTTTAGTC CCTTGGGCTC TGCTACCCTC CTGCTGGAGG TGAGAGCATC    60
CTGTGTGCAA CCAGAGATGC CCTCTGGCTT TCAGACCTGC CTGCTTTTCA CCCTCAGCCC   120
TTTCTCACTC AGCAAAATTG TGGGGGTCCC TAGTCAGCAG CTCCCTGGGC AGCTCTCTGA   180
GCAAGGTGGT CTCTGTGGTC ATGAAGGAGA GCCGGCTAGG ACAGTGCCGG AAACTCAGCT   240
GCCTCTCCCC TTCAACTCAG CTGGCCCCCC GCACCTGAAG TGCACAGGAG CCGGGAAGAG   300
AGTCTGGAGC CCACCCCGGA GGGCAGCACA GGAGGTGTCT CTGCAGCTGG TGTCCTGCCA   360
CCCCTGCAGG CAGCACACGT CCCGGGCATT CTCCTTAGCC ACAGACAGAA CAGCCAGTGC   420
CAGAGTCTGC TGTCGTTCCC CTTTAAGCAC ACTCATTCAC CACACCCGAG GAGGCCAGAG   480
GTGCAGGGAG CATGGGCTGT CGCTTCCCCT TTAAGCACAC TCATTCACCA CACCCGAGGA   540
GGCCAGAAGT GCAGGGAGCA TGGGCTGGGT GCACCTCCGC AGGAGAGAAG GCTGAGCCAC   600
CGCCGTCCCG GGAGCCCGGC TCCCAGGCCT CTCGTTTTCC CCTACCTCCC TAAGACTTTT   660
CTGTCACTCT CTGGCCATTG AAAGGCTTCT GTTCCTTAAA GTGCTGTTAC ACTCTCCTTT   720
CCCAGGATGC AGCAAGCCAA AACAGTACCA CTGCACGTCA GCCTGGGTGA CAGAGTGAGA   780
CCCTATCTTA AAAAAAAAAA AAAAAA                                        806
```

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 696 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

```
GAGTTCCCNA CGCGGTGGCG NCCGTTTTAG AAATTAGTGG ATCCCCCCGG GCTGGCAGGG    60
AATTCGGCAC GAGCACAGAG GAAAGCGGGT GCCCGGCATG GCCATCCTGA TGTTGCTGGC   120
GGGATCCCCA TGCACCTTGT CCTTCTCCAC TGATACTGGC AGCTCGGCTC CTGGACCCAA   180
GATCCCTTGA GTGGAATTCT GCAGTGCAAG AGCCCTTCGT GGGAGCTGTC CCATGTTTCC   240
ATGGTCCCCA GTCTCCCCTC CACTTGGTGG GGTCACCAAC TACTCACCAG AAGGGGGCTT   300
ACCAAGAAAG CCCTAAAAAG CTGTTGACTT ATCTGCGCTT GTTCCAACTC TTATGCCCCC   360
AACCTGCCCT ACCACCACCA CGCGCTCAGC CTGATGTGTT TACATGGTAC TGTATGTATG   420
GGAGAGCAGA CTGCACCCTC CAGCAACAAC AGATGAAAGC CAGTGAGCCT ACTAACCGTG   480
```

CCATCTTGCA AACTACACTT TAAAAAAAAC TCATTGCTTT GTATTGTAGT AACCAATATG 540

TGCAGTATAC GTTGAATGTA TATGAACATA CTTTCCTATT TCTGTTCTTT GAAAATGTCA 600

GAAATATTTT TTTCTTTCTC ATTTTATGTT GAACTAAAAA GGATTAAAAA AAAAATCTCC 660

AGAMAAAAAA AAAAAAAAAA AAATTACTGC GGTCCG 696

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1007 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

AATTCGGCAC GAGGAAAAAA TACCATTTGT GTATGATACC CAATTTGGAT CTCAATTTGG 60

ATAGAGATTT GGTGCTTCCA GATGTRAGTT ATCAGGTGGA ATCCAGTGAG GAGGATCAGT 120

CTCAGACTAT GGATCCTCAA GGACAAACTC TGCTGCTTTT TCTCTTTGTG GATTTCCACA 180

GTGCATTTCC AGTCCAGCAA ATGGAAATCT GGGGAGTCTA TACTTTGCTC ACAACTCATC 240

TCAATGCCAT CCTTGTGGAG AGCCACAGTG TAGTGCAAGG TTCCATCCAA TTCACTGTGG 300

ACAAGGTCTT GGAGCAACAT CACCAGGCTG CCAAGGCTCA GCAGAAACTA CAGGCCTCAC 360

TCTCAGTGGC TGTGAACTCC ATCATGAGTA TTCTGACTGG AAGCACTAGG AGCAGCTTCC 420

GAAAGATGTG TCTCCAGACC CTTCAAGCAG CTGACACACA AGAGTTCAGG ACCAAACTGC 480

ACAAAGTATT TCGTGAGATC ACCCAACACC AATTTCTTCA CCACTGCTCA TGTGAGGTGA 540

AGCAGCTAAC CCTAGAAAAA AAGGACTCAG CCCAGGGCAC TGAGGACGCA CCTGATAACA 600

GCAGCCTGGA GCTCCTAGCA GATACCAGCG GGCAAGCAGA AAACAAGAGG CTCAAGAGGG 660

GCAGCCCCCG CATAGAGGAG ATGCGAGCTC TGCGCTCTGC CAGGGCCCCG AGCCCGTCAG 720

AGGCCGCCCC GCGCCGCCCG GAAGCCACCG CGGCCCCCCT CACTCCTAGA GGAAGGGAGC 780

ACCGCGAGGC TCACGGCAGG GCCCTGGCGC CGGGCAGGGC GAGCCTCGGA AGCCGCCTGG 840

AGGACGTGCT GTGGCTGCAG GAGGTCTCCA ACCTGTCAGA GTGGCTGAGT CCCAGCCCTG 900

GGCCCTGAGC CGGGTCCCCT TNCGCAAGCG CCCACCGATC CGGARGCTGC GGGCAGCCGT 960

TATCCCGTGG TTTAATAAAG TGCCGCGCGC TCACCAAAAA AAAAAAA 1007

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2017 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

```
AATTCGGCAC GAGCGGATCC GTTGCGGCTG CAGCTCTGCA GTCGGGCCGT TCCTTCGCCG     60
CCGCCAGGGG TAGCGGTGTA GCTGCGCACG TCGCGCGCGC TACCGCACCC AGGTTCGGCC    120
CGTAGCGTCT GGCAGCCCGG CGCCATCTTC ATCGAGCGCC ATGGCCGCAG CCTGCGGGCC    180
GGGAGCGGCG GGTACTGCTT GCTCCTCGGC TTGCATTTGT TTCTGCTGAC CGCGGGCCCT    240
GCCTGGGCTG GAACGACCCT GACAGAATGT TGCTGCGGGA TGTAAAAGCT CTTACCCTCC    300
ACTATGACCG CTATACCACC TCCCGCAGCT GGATCCCATC CCACAGTTGA AATGTGTTGG    360
AGGCACAGCT GGTTGTGATT CTTATACCCC AAAAGTCATA CAGTGTCAGA ACAAAGGCTG    420
GGATGGGTAT GATGTACAGT GGGAATGTAA GACGGACTTA GATATTGCAT ACAAATTTGG    480
AAAAACTGTG GTGAGCTGTG AAGGCTATGA GTCCTCTGAA GACCAGTATG TACTAAGAGG    540
TTCTTGTGGC TTGGAGTATA ATTTAGATTA TACAGAACTT GGCCTGCAGA AACTGAAGGA    600
GTCTGGAAAG CAGCACGGCT TTGCCTCTTT CTCTGATTAT TATTATAAGT GGTCCTCGGC    660
GGATTCCTGT AACATGAGTG GATTGATTAC CATCGTGGTA CTCCTTGGGA TCGCCTTTGT    720
AGTCTATAAG CTGTTCCTGA GTGACGGGCA GTATTCTCCT CCACCGTACT CTGAGTATCC    780
TCCATTTTCC CACCGTTACC AGAGATTCAC CAACTCAGCA GGACCTCCTC CCCCAGGCTT    840
TAAGTCTGAG TTCACAGGAC CACAGAATAC TGGCCATGGT GCAACTTCTG GTTTTGGCAG    900
TGCTTTTACA GGACAACAAG GATATGAAAA TTCAGGACCA GGGTTCTGGA CAGGCTTGGG    960
AACTGGTGGA ATACTAGGAT ATTTGTTTGG CAGCAATAGA GCGGCAACAC CCTTCTCAGA   1020
CTCGTGGTAC TACCCGTCCT ATCCTCCCTC CTACCCTGGC ACGTGGAATA GGGCTTACTC   1080
ACCCCTTCAT GGAGGCTCGG GCAGCTATTC GGTATGTTCA AACTCAGACA CGAAAACCAG   1140
AACTGCATCA GGATATGGTG GTACCAGGAG ACGATAAAGT AGAAAGTTGG AGTCAAACAC   1200
TGGATGCAGA AATTTTGGAT TTTTCATCAC TTTCTCTTTA GAAAAAAAGT ACTACCTGTT   1260
AACAATTGGG AAAAGGGGAT ATTCAAAAGT TCTGTGGTGT TATGTCCAGT GTAGCTTTTT   1320
GTATTCTATT ATTTGAGGCT AAAAGTTGAT GTGTGACAAA ATACTTATGT GTTGTATGTC   1380
AGTGTAACAT GCAGATGTAT ATTGCAGTTT TTGAAAGTGA TCATTACTGT GGAATGCTAA   1440
AAATACATTA ATTTCTAAAA CCTGTGATGC CCTAAGAAGC ATTAAGAATG AAGGTGTTGT   1500
ACTAATAGAA ACTAAGTACA GAAAATTTCA GTTTTAGGTG GTTGTAGCTG ATGAGTTATT   1560
```

ACCTCATAGA GACTATAATA TTCTATTTGG TATTATATTA TTTGATGTTT GCTGTTCTTC 1620

AAACATTTAA ATCAAGCTTT GGACTAATTA TGCTAATTTG TGAGTTCTGA TCACTTTTGA 1680

GCTCTGAAGC TTTGAATCAT TCAGTGGTGG AGATGGCCTT CTGGTAACTG AATATTACCT 1740

TCTGTAGGAA AAGGTGGAAA ATAAGCATCT AGAAGGTTGT TGTGAATGAC TCTGTGCTGG 1800

CAAAAATGCT TGAAACCTCT ATATTTCTTT CGTTCATAAG AGGTAAAGGT CAAATTTTTC 1860

AACAAAAGTC TTTTAATAAC AAAAGCATGC AGTTCTCTGT GAAATCTCAA ATATTGTTGT 1920

AATAGTCTGT TTCAATCTTA AAAAGAATCA ATAAAAACAA ACAAGGGAAA AAAAAAAAAA 1980

AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAA 2017

(2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 699 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

GNGTTTTTTC CAGCCAGGAA GTGACCGNTA CTGCAGCACG AGANAGATTG GTTGGGTTGG 60

TTGRAAATGA CYCTGAACAT TTATTTCCAT TGCAATTTCT GTGGCTGAGG AGACTTAAAC 120

TTTACAAGTA TTATCCTTTT AAGATCATTT TAATTTTAGT TGAGTGCAGA GGGCTTTTAT 180

AACAAACGTG CAGAAATTTT GGAGGGCTGT GATTTTTCCA GTATTAAACA TGCATGCATT 240

AATCTTGCAG TTTATTTTCT CATTGTGTAT GTATATATCG CTTTTCTCTG CAGCACGATT 300

TCTCTTTTGA TAAWKCCCTT TAGGGCACAA CTAGTTATCA GTAACTGAAT GTATCTTAAT 360

CATTATGGCT GCTTCTGTTT TTTCATTAAC AAAGGTTATT CATATGTTAG CATATAGTTT 420

CTTTGCACCC ACTATTTATG TCTGAATCAT TTGTCACAAG AGAGTGTGTG CTGATGAGAT 480

TGTAAGTTTG TGTGTTTAAA CTTTTTTTTG AGCGAGGGAA GAAAAAGCTG TATGCATTTC 540

ATTGCTGTCT ACAGGTTTCT TTCAGATTAT GTTCATGGGT TTGTGTGTAT ACAATATGAA 600

GAATGATCTG AAGTAATTGT GCTGTATTTA TGTTTATTCA CCAGTCTTTG ATTAAATAAA 660

AAGGAAAACC AGAAAAAAAA AAAAAAAAAA AAAAAAAAA 699

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1264 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

```
GGCACGAGGG CACTGTTTCC TCAGTCCATG GCTGAGTACA TCACCGGTGT TTTCTCTCTT      60
ATTCCTCCCA TCAAGCCTAA AAGGAATCTC TATTGGAGAT ACTGCCATTA GTGTTCCTTT     120
TATAGGTGAG GAACTGAGGC ATAKAGGGTT CCCCAGTTGA ACCAACTGAT AAATAGTAGA     180
ACTTGGATTT TAATTCAGTC TTGATGCCAG GGATAAGGCT CTTACTTTCT ACCTTAGGCT     240
ATTTCTAGGA AACGCAGGAG AGTGTTGAAG GGGCAGAGAA AGGGATCCAG TTCCTTTCTG     300
TCCCGCATCC TAGTCCCTGA GAAGCAAAGA ARAATGTGTG GCTTCTTTTG CTTTGCTTTT     360
GTTGTCATCC CACACATCTC CAGGGGAMCT GGGCTCTTGA TCTTGGSCTC TTCCCCTTTA     420
ACTGTTAAGT GGGAGCARGT AAGGGGGTAC AGTAGGGCTG GCCTGGAGTT AGAGGCTTGG     480
ATGCCTTAGC TCCTCTGTCT GCACTCCAGA ACTGCCTGAC TTCATTTCGT ATGTTGTCCT     540
TTGTTTTGAC AATTGATCCA TGTCCCAGTC CGTCTCTTCT TCCTTCTTGA TACTTACACT     600
GCTTCTTTCT GTTGGTTTCC AGTGTTTAAC ACTGTATACA ACAGTGACGA CAACGTGTTT     660
GTGGGGGCCC CCACGGGCAG CGGGAAGACT ATTTGTGCAG AGTTTGCCAT CCTGCGAATG     720
CTGCTGCAGA GCTCGGAGGG GCGCTGTGYS TWCWTCACCM CCATGGAGGC CCTGGCCAGA     780
RCAGGTATGA CGTGGCGCTG TGTCATGTGA ATTTCCCAAG AAGCATTTCA TCTGTGATTC     840
CGTATGAAGG CTTTCTAAGC CCTGAAATTT GCAGGGTCAT TTCCTCAGTT TGTGTATTAA     900
AGAAAAGCTG CCCCAGCCAA GCGTGGTGGC TCACGCCTGT AATCCCAGCA CTTTGGGAGG     960
CCGAGGCGGG CAGATCTCCG GAGATCAGGA GTTCGAGACC AGCCTGGCCA ACATGGTGRA    1020
ACCCTGTCTC TACTAAAAWT ACAGAAATTA GCTGGGNGTG GTGGTGTGCG CCTGTAATCC    1080
CAGCTACTTG GAAGGCTGAG GCAGGAGAAT CGCTTGAACC CGGGAGGCGG AGGTTGCAGT    1140
GAGCCAAGTT CGCACCACTG CACTCCAGCC TGGGCAACAA GAGCGAGACT TCATCTCAAA    1200
AAAAAAAAAA AAAAACTCGA GGGGGGGCCC GGTACCCAAT TCGCCCTATA GTGATCGTAT    1260
TACA                                                                 1264
```

(2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 997 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

```
ATTGGAAGTT GTTTTGCAAC CTGGGCTTTT ATACAGAAGA ATACGAATCA CAGGTGTGTG    60
AGCATCTACT TAATTAATTT GCTTACAGCC GATTTCCTGC TTACTCTGGC ATTACCAGTG   120
AAAATTGTTG TTGACTTGGG TGTGGCACCT TGGAAGCTGA AGATATTCCA CTGCCAAGTA   180
ACAGCCTGCC TCATCTATAT CAATATGTAT TTATCAATTA TCTTCTTAGC ATTTGTCAGC   240
ATTGACCGCT GTCTTCAGCT GACACACAGC TGCAAGATCT ACCGAATACA AGAACCCGGA   300
TTTGCCAAAA TGATATCAAC CGTTGTGTGG CTAATGGTCC TTCTTATAAT GGTGCCAAAT   360
ATGATGATTC CCATCAAAGA CATCAAGGAA AAGTCAAATG TGGGTTGTAT GGAGTTTAAA   420
AAGGAATTTG GAAGAAATTG GCATTTGCTG ACAAATTTCA TATGTGTAGC AATATTTTTA   480
AATTTCTCAG CCATCATTTT AATATCCAAT TGCCTTGTAA TTCGACAGCT CTACAGAAAC   540
AAAGATAATG AAAATTACCC AAATGTGAAA AAGGCTCTCA TCAACATACT TTTAGTGACC   600
ACGGGCTACA TCATATGCTT TGTTCCTTAC CACATTGTCC GAATCCCGTA TACCCTCAGC   660
CAGACAGAAG TCATAACTGA TTGCTCAACC AGGATTTCAC TCTTCAAAGC CAAAGAGGCT   720
ACACTGCTCC TGGCTGTGTC GAACCTGTGC TTTGATCCTA TCCTGTACTA TCACCTCTCA   780
AAAGCATTCC GCTCAAAGGT CACTGAGACT TTTGCCTCMC CTAAAGAGAC CAAGGTYAGA   840
AAGAAAAATT AAGANGTGGA AATAATGGCT AAAAGACAGG NTTTTTGTGG TACCAATTCT   900
GGGCTTTATG GGACCNTAAA GTTATTATAG CTTGGAAGGT AAAAAAAAAA AAAGGGNGGG   960
CGCTCTAGAG GTTCCCCGAG GGGCCAGCTT AGGGTGC                            997
```

(2) INFORMATION FOR SEQ ID NO: 34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1914 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

```
GTGTGAGAGG CCTCTCTGGA AGTTGTCCCG GGTGTTCGCC GCTGGAGCCC GGGTCGAGAG    60
GACGAGGTGC CGCTGCCTGG AGAATCCTCC GCTGCCGTCG GCTCCCGGAG CCCAGCCCTT   120
TCCTAACCCA ACCCAACCTA GCCCAGTCCC AGCCGCCAGC GCCTGTCCCT GTCACGGACC   180
CCAGCGTTAC CATGCATCCT GCCGTCTTCC TATCCTTACC CGACCTCAGA TGCTCCCTTC   240
TGCTCCTGGT AACTTGGGTT TTTACTCCTG TAACAACTGA AATAACAAGT CTTGATACAG   300
```

```
AGAATATAGA TGAAATTTTA AACAATGCTG ATGTTGCTTT AGTAAATTTT TATGCTGACT    360
GGTGTCGTTT CAGTCAGATG TTGCATCCAA TTTTTGAGGA AGCTTCCGAT GTCATTAAGG    420
AAGAATTTCC AAATGAAAAT CAAGTAGTGT TTGCCAGAGT TGATTGTGAT CAGCACTCTG    480
ACATAGCCCA GAGATACAGG ATAAGCAAAT ACCCAACCCT CAAATTGTTT CGTAATGGGA    540
TGATGATGAA GAGAGAATAC AGGGGTCAGC GATCAGTGAA AGCATTGGCA GATTACATCA    600
GGCAACAAAA AAGTGACCCC ATTCAAGAAA TTCGGGACTT AGCAGAAATC ACCACTCTTG    660
ATCGCAGCAA AAGAAATATC ATTGGATATT TTGAGCAAAA GGACTCGGAC AACTATAGAG    720
TTTTTGAACG AGTAGCGAAT ATTTTGCATG ATGACTGTGC CTTTCTTTCT GCATTTGGGG    780
ATGTTTCAAA ACCGGAAAGA TATAGTGGCG ACAACATAAT CTACAAACCA CCAGGGCATT    840
CTGCTCCGGA TATGGTGTAC TTGGGAGCTA TGACAAATTT TGATGTGACT TACAATTGGA    900
TTCAAGATAA ATGTGTTCCT CTTGTCCGAG AAATAACATT TGAAAATGGA GAGGAATTGA    960
CAGAAGAAGG ACTGCCTTTT CTCATACTCT TTCACATGAA AGAAGATACA GAAAGTTTAG   1020
AAATATTCCA GAATGAAGTA GCTCGGCAAT TAATAAGTGA AAAAGGTACA ATAAACTTTT   1080
TACATGCCGA TTGTGACAAA TTTAGACATC CTCTTCTGCA CATACAGAAA ACTCCAGCAG   1140
ATTGTCCTGT AATCGCTATT GACAGCTTTA GGCATATGTA TGTGTTTGGA GACTTCAAAG   1200
ATGTATTAAT TCCTGGAAAA CTCAAGCAAT TCGTATTTGA CTTACATTCT GGAAAACTGC   1260
ACAGAGAATT CCATCATGGA CCTGACCCAA CTGATACAGC CCCAGGAGAG CAAGCCCAAG   1320
ATGTAGCAAG CAGTCCACCT GAGAGCTCCT TCCAGAAACT AGCACCCAGT GAATATAGGT   1380
ATACTCTATT GAGGGATCGA GATGAGCTTT AAAAACTTGA AAAACAGTTT GTAAGCCTTT   1440
CAACAGCAGC ATCAACCTAC GTGGTGGAAA TAGTAAACCT ATATTTTCAT AATTCTATGT   1500
GTATTTTTAT TTTGAATAAA CAGAAAGAAA TTTTGGGTTT TTAATTTTTT TCTCCCCGAC   1560
TCAAAATGCA TTGTCATTTA ATATAGTAGC CTCTTAAAAA AAAAAAAAAC CTGCTAGGAT   1620
TTAAAAATAA AAATCAGAGG CCTATCTCCA CTTTAAATCT GTCCTGTAAA AGTTTTATAA   1680
ATCAAATGAA AGGTGACATT GCCAGAAACT TACCATTAAC TTGCACTACT AGGGTAGGGA   1740
GGACTTAGGG ATGTTTCCTG TGTCGTATGT GCTTTTCTTT CTTTCATATG ATCAATTCTG   1800
TTGGTATTTT CAGTATCTCA TTTCTCAAAG CTAAAGAGAT ATACATTCTG GATACTTGGG   1860
AGGGGAATAA ATTAAAGTTT TCACACTGNA AAAAAAAAAA AAAAAAAAAC TCGA         1914
```

(2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1020 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

```
CCNTTNNTTT TTTTTTTTTG CAAGACAAAA TATACTTTAT TGTGACAGCA AATGCACATA      60
GTGCTGTAGG TAAGGCATGC TACTAGGAAT CTGCATATAA TCAAAAGCCA GTATGGAAAT     120
GAATGGAAAT GAATGCTGTT GTTCTCAGAT TGAGTCCATG GTGGAGAAAG GATAGTTTGT     180
GTCCACTTAT TTCAAATGCA GTATCATACC TACTTAATCA GTTACCTATG CTTCTAACCA     240
ACAGCCCAGT GGCAAATAGG AGGAACTTAA CTGTACTCAG AAGTCACTTT TAATATCAAC     300
GACAGAAATA TTTCACTAAT TCAACTGAGG CAAATTTCCT TTCTAGACAA AGGACCTAGA     360
AATTGAGCAT GCAAAACATC CATCCATTCA TTCATTCAAA TAATTAGCCA ATTTTACCGT     420
CATTTAATTC CACCAGAAGC AAATACTAGA ATATCTAGAA GTAGTTTGGG TAAAGAAACA     480
TTTACATTTT AATATTGTGT AATGTCATAA ATTTGGGGCT AAAATAACAC CAGGTCAAAT     540
TTGATCCCTT TGTATGTGAG GGTACAAAGT ACAGTTTTCG TTTCAACAGC TGAACTTCTG     600
AGAGAAGAGC TGAAAAAAAT GCTAAATAAG AGATCTAGGC CTTTGATGGA AACTATTAGG     660
CTCTACAGAC TTGTCAAAAA ATCAATGCAA AACTGAGGGG GAAAGGCTGA AATGCTTTGT     720
AAAGCAGTAT TTTTAGACAA GTTGCTTCAT TTCCCCCTTT TCTAAAACAG ATGCAGATTA     780
AATGTTTTTT TGCATGAATG CACATTGACA TTCTGTTCAA CTGTTTTCTA AATGCAACAC     840
TGCGGGTTTC AACAGTATGC TTTCATTTAA ACAAAGAATA TTATATGCAT GGTCAATTTA     900
GTTTAAGAGA TGAAAAAAAA CTTTACTACT ATGAAAATTG CTTATCAAAT ACTCTCCTCT     960
TTTATAAGGT GTTTTTARGC AACACAGGAC CGGTNGAACC GANCAAATTT ATAATTATAC    1020
```

(2) INFORMATION FOR SEQ ID NO: 36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 781 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

```
AACTCCTGAC CTCAAGTGCT CCACCTGCGT TGGCTTCCCA AAGTGCTGGG ATACAGGAGT      60
RAGCCACTGC GCCTGGCTGA TCCCAGCACT TTTMAAATGA TGCCGCTCAA AGCCGTGACT     120
TGGCCTACTT TGAACAGCAA ACTTGTTGCT GCTGTTGTCA ACCTGAAGGC CTCTCAAATG     180
```

CCAGCTTCAA GCAGGGTGTG AATTGGCCAG TGTCAGATCT CAGGAGTCCT GTGTTGAGAG 240

TGTGGCTTTC AGCTGCGGGG AGCTGCACTT GGTGGGGAAA GCCAGGCAGG TCACCCTCAC 300

AGCCAGATAA TGTGGAGGTC AGAACCCAAG GAAGGGAGTG AGACCTCCAC TCCCAGTGGG 360

GGACCTGGCC ACCCATCCTT GGGGACCTGA GAAAGCGTAC TTCACCTTGG GGTGAAGGCT 420

GGGTGGGGCC AGAGGGACCA GTGCCCTCCT CAGTGCTTAG GGGCAGAGCC ACCTGCAGCA 480

ATGGTATCTG CATATTAGCC CCTCTCCACC TTCTTTCTCC CGCTGAATCA TTTCCCTCAA 540

AGCCCAAGAG CTGTCACTGC TTCTTTCTCC CTGGGAAGAA TGCGTGGACT CTGCCTGGTG 600

ATAGACTGAA GCCAGAACAG TGCCACACCC TCGCCTTAAT TCCTTGCTAG GTGTTCTCAG 660

ATTTATGAGA CTTCTTAGTC AAATATGAGG GAGGTTGGAT GTGGTGGCTT GTGCCTGTAA 720

TCCCAGCATT TTGGGAAGCC GAGGTGGGAG GATCCCTTGA AGCCAGGAGT TTGAGACAAG 780

C 781

(2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 966 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

GGCACGAGGA AGCAGCTGGG GGCTGATCAG GGGGAGCACG CAGCCCTCCG ATTGCAGGGC 60

TGCCTATTTG AGTGGCAGCT CCTCTTGAAA CAATGCAGAA CAAGCCCAGG GCCCCACAGA 120

AAAGGGCACT GCCCTTCCCA GAACTTGAGC TCCGGGACTA CGCATCTGTT CTCACCAGAT 180

ACAGCTTGGG GCTGAGGAAC AAAGAGCCTT CCCTGGGCCA CAGGTGGGGG ACCCAGAAGC 240

TGGGCAGGAG CCCCTGTTCT GAAGGGTCCC AGGGCCACAC CACAGATGCT GCTGACGTGC 300

AGAACCACTC TAAAGAAGAA CAGAGAGACG CAGGAGCACA GAGGARGTGC GGCCAGGGGA 360

GGCACACCTG GGCGTACAGG NGAGGGGCGC AGGACACTTC GAGGCTGACA GGAGACCCAC 420

GTGGTGGGGA AAGGAGCCCC CCAAAGTGTC AGAGCATGAA GCAGCAGGAA GGAGCTCCCT 480

CGGGCCACTG CTGGGATCAG TGGTGCCATG GAGCAAGCGA GGTTGTTTGG CCTGAAAGCC 540

GGAAGCGTGC CCAAATCTTT SCATCACCAT GTAGGCAGTC ACCTCGCTCC TCAGCACTCG 600

GGGCAGGACA GAAGCTTGCT GTCTGCTCAC CAGACATCCT GTGCTGCCCT ACAGACACCT 660

TGCTCGCCAG CCATCCCCAC TCACTTCTGA CCGGGACCCA ATTCTCTGGC CAAACCCAGG 720

CTCTAGCACC GTCTTGGTGT GCTTGAGAAA CATCTAGTTT AAGTCAAAAT CCAATGTCTT 780

TTTAATATAT AGACTATATG TACCTATGGA CTAGAGGTGA ATATATATAC ATCATATCAA 840

ATTCAAGTGA CCCAGTATTT CGGGAGAACC CACTATGTCC CCAGCCTGCA TGGGAAGCTG 900

GGGATTCTGG CATGAACTGC ACCTTATCTT CCTCGAGGGG GGGCCGGTAC CAATTGCCNA 960

TAGTGG 966

(2) INFORMATION FOR SEQ ID NO: 38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 416 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

GAATTCGGCA CGAGGTAATA GGAGCCCTCG TACCTCTTGT GTTCCTTACA AACATTCTCA 60

TCAGTAGCTC TACGCGTTGA CTGGGTGGTT TGARATGGCT GGTATACACA GGGCTTTCTT 120

GGTGTTCTGT CTCTGGGGCT TARCTTTGTG TGTGGTTGGA GGGCCCTGGT GAGATTGGAA 180

GTACCAGAGA GTGCTGTGTC AGGGGCAGAG GGGCCTGTCG CTGGAGCTGG AGGGTGCCTG 240

CCTTTGTGTC TGACTCARTC TCCTGTCTGC CTTGCCCCCT CAGGGTCTCG CCAGCCCAGC 300

CTCTGTGGGA ATCTAAAAGG ARTGGATGTG GACGTKTGAC CAAGCACATC TCAGCTTTTA 360

ATACCTGGGC TATTTATAGA CCTTTGGGGG GAATNGCTTG TGGAACAACA AGGGTT 416

(2) INFORMATION FOR SEQ ID NO: 39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1114 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

TGTGTATTTG GGGGGACTGA AGGGTACGTG GGGCGAAACA AAACCGGCCA TGGCAGCAGC 60

GGAGGAGGAG GACGGGGGCC CCGAAGCCAA AATCGCGAGC GGGGCGGGGC GGGCGCGACC 120

TTCGAATGTA ATATATGTTT GGAGACTGCT CGGGAAGCTG TGGTCAGTGT GTGTGGCCAC 180

CTGTACTGTT GGCCATGTCT TCATCAGTGG CTGGAGACAC GGCCAGAACG GCAAGAGTGT 240

CCAGTATGTA AAGCTGGGAT CAGCAGAGAG AAGGTTGTCC CGCTTTATGG GCGAGGGAGC 300

CAGAAGCCCC AGGATCCCAG ATTAAAAACT CCACCCCGCC CCCAGGGCCA GAGACCAGCT 360

CCGGAGAGCA GAGGGGGATT CCAGCCATTT GGTGATACCG GGGGCTTCCA CTTCTCATTT 420

GGTGTTGGTG CTTTTCCCTT TGGCTTTTTC ACCACCGTCT TCAATGCCCA TGAGCCTTTC 480

CGCCGGGGTA CAGGTGTGGA TCTGGGACAG GGTCACCCAG CCTCCAGCTG GCAGGATTCC 540

CTCTTCCTGT TTCTCGCCAT CTTCTTCTTT TTTTGGCTGC TCAGTATTTG AGCTATGTCT 600

GCTTCCTGCC CACCTCCAGC CAGAGAAGAA TCAGTATTGA GGGTCCCTGC TGACCCTTCC 660

GTACTCCTGG ACCCCCTTGA CCCCTCTATT TCTGTTGGCT AAGGCCAGCC CTGGACATTG 720

TCCAGGAAGG CCTGGGGAGG AGGAGTGAAG TCTGTGCATA GATGGGAGAG CCTTCTGCTC 780

AGAGGCTCAC TCAGTAACGT TGTTTAATTC TCTGCCCTGG GGAAGGAGGA TGGATTGAGA 840

GAATGTCTTT CTCCTCTCCT AAGTCTTTGC TTTCCCTGAT TTCTTGATTT GATCTTCAAA 900

GGTGGGCAAA GTTCCCTCTG ACTCTTCCCC CACTCCCCAT CTTACTGATT TAATTTAATT 960

TTTCACTCCC CAGAGTCTAA TATGGATTCT GACTCTTAAG TGCTTCCGCC CCCTCACTAC 1020

CTCCTTTAAT ACAAATTCAA TAAAAAAGGT GAAATATAAA AAAAAAAAAA AAAAAACYCG 1080

GGGGGGGCCC CGGTCCCCAT TCCCTTTGGG GGGT 1114

(2) INFORMATION FOR SEQ ID NO: 40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 602 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

GGGTCGACCC ACGCGTCCGT CCCAGGCCAC AAGACATTTC CTGCTCGGAA CCTTGTTTAC 60

TAATTGTCTC TGTGGCACAT TTTGTTTCCC GTGCCTTGGG TGTCAAGTTG CAGCTGATAT 120

GAATGAATGC TGTCTGTGTG GAACAAGCGT CGCAATGAGG ACTCTCTACA GGACCCGATA 180

TGGCATCCCT GGATCTATTT GTGATGACTA TATGGCAACT CTTTGCTGTC CTCATTGTAC 240

TCTTTGCCAA ATCAAGAGAG ATATCAACAG AAGGAGAGCC ATGCGTACTT TCTAAAAACT 300

GATGGTGAAA AGCTCTTACC GAAGCAACAA AATTCAGCAG ACACCTCTTC AGCTTGAGTT 360

CTTCACCATC TTTTGCAACT GAAATATGAT GGATATGCTT AAGTACAACT GATGGCATGA 420

AAAAAATCAA ATTTTTGATT TATTATAAAT GAATGTTGTC CCTGAACTTA GCTAAATGGT 480

GCAACTTAGT TTCTCCTTGC TTTCATATTA TCGAATTTCC TGGCTTATAA ACTTTTTAAA 540

TTACATTTGA AATATAAACC AAATGAAATA TTTTACTGAA AAAAAAAAAA AAAAAANCCC 600

CA 602

(2) INFORMATION FOR SEQ ID NO: 41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 970 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

GGCAGAGCTT AGGAGAACAG CTCCCTTTGG ATCCCTNTCA AAGGTGATAC CATTGGCTCC 60

CAGCTTAGAG TAAGAAGCTC TGAGAAGTTG AATGAAGGGT GAGATAGAGA TGCTGAACCC 120

ATTCTTSCAG CTTCTTCTAG TGTTGTTATT TCCAGAATGG CCAACACCCC TACATTGATA 180

CATAAACACA TTCCAAGGCC TTGTGTAATA CAAAGTTCAC CGTCCTCCTG GAATAGGAGC 240

CCTGGGTTCT AGTTCTCACT CTGCCACTGG GGGAAAATCC AATTAAAGTC TGGTTTAGTC 300

AGCTTGGGTC ACCATAGACT GGGTGGCTTA AACAGCAGAC ATTTATTTCT GGTAGTTTCT 360

GGAGGCTACA AATCTAAGAG CAAGGTGCCA GCATGGTCAC ATTCTGGTGA GGGSCCTCTT 420

CCTGGCTTGT AGACGGCTGC YTTCTCACCG TGTGCTCACA TAGCCTTTCG TGTGTGTGTG 480

TGTGTGTGTG TGCGTKCGTG CAAGCTTCCK GATGTCTCTT CTTAGAAGGA CACCAACCCC 540

ATCATGAGAG CCCTACTCTC ATGACTTAGC CTAACCCTAA TTACCCTCCA AAGGCCCCAT 600

CTCCAAATGC CATCACATTG GAGGGTAGAG CTTCAACATA GGGATTTTGG GGGACACAAA 660

CATTCAGTCC ATAACAAAGG CTGTAGTCCT TARTTTCCTT GTCTGTGAAA TGAGAGTGTT 720

GAGATTCTTT CTAGCCTTTA TCATTTATAA TTCTGTGAGA TGTAGATTTG CATTATTTTC 780

GAGTTCGAGT TATATGAAAT GTTTCCCTCT ACATTTTCTT GGGCAACTGA GAACTGAATA 840

GGGCTAGGTT TAAATAGAGT TAGGCAGTTA GGCTTATTCT TTTATTTAAT AAGCATTTTT 900

GGAGCATCTA CGGTGTTCCA GGAACTGAAC TGTTGTAAAC ATTGGAGCTG TAACAGAGAA 960

CAAAAGAGAC 970

(2) INFORMATION FOR SEQ ID NO: 42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1002 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

GAATTCGGCA CGAGCCGAGG TCGGCAGCAC AGAGCTCTGG AGATGAAGAC CCTGTTCCTG 60
GGTGTCACGC TCGGMCTGGC CGCTGCCCTG TCCTTMACCC TGGRGGAGGA GGATATCACA 120
GGGACCTGGT ACGTGAAGGC CATGGTGGTC GATAAGACTT TCCGGAGACA GGAGGCCCAG 180
AAGGTGTCCC CAGTGAAGGT GACAGCCCTG GGCGGTGGGA AGTTGGAAGC CACGTTCACC 240
TTCATGAGGG AGGATCGGTG CATCCAGAAG AAAATCCTGR TGCGGAAGAC GGAGGAGCCT 300
GGCAAATACA GCGCCTGTGA GCCCCTCCCC CAYTCCCACC CCCACCYTCC CCCACCGCCA 360
ACCCCAGTGC ACCAGCCTCC ACAGGTAGAG AGTGCCCAGG CTGCCCTTTT GCCAGGGCCC 420
CAGCTCTGCC CACCTCCAAG GAGGGGCTGG CCTCTCCTTC CTGGGGGGCT GGTGGCCCTG 480
ACATCAGACA CCGGGTGTGA CAGGCTTGTC CGCAGTCGAG ATGGACCAGA TCACGCCTGC 540
CCTCTGGGAG GCCCTAGCCA TTGACACATT GAGGAAGCTG AGGATTGGGA CAAGGAGGCC 600
AAGGATTAGA TGGGGGCAGG AAGCTCATGT ACCTGCAGGA GCTGCCCAGG AGGGACCAYT 660
ACATCTTTTA CTGCAAAGAC CAGCACCATG GGGGCSTGCT CCACATGGGA AAGCTTGTGG 720
GTAGGAATTC TGATACCAAC CGGGAGGCCC TGGAAGAATT TAAGAAATTG GTGCAGCGCA 780
AGGGACTCTC GGAGGAGGAC ATTTTCACGC CCCTGCAGAC GGGAAGCTGC GTTCCCGAAC 840
ACTAGGCAGC CCCCGGGTCT GCACCTCCAG AGCCCACCCT ACCACCAGAC ACAGAGCCCG 900
GACCACCTGG ACCTACCCTC CAGCCATGAC CCTTCCCTGC TCCCACCCAC CTGACTCCAA 960
ATAAAGTCCT TCTCCCCCAA AAAAAAAAAA AAAAAAACTC GA 1002

(2) INFORMATION FOR SEQ ID NO: 43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2581 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

TGCAAAACCA CTGGACACTG GACAAGTACG GGATCCTGGS CGACGCACGC CTCTTCTTTG 60
GGCCCCAGCA CCGGSCCGTC ATCCTTCGGT TGTCCAACCG CCGCGCACTG CGCCTCCGTG 120
CCAGCTTCTC CCAGCCCCTC TTCCAGGCTG TGGSTGCCAT CTGCCGCCTC CTCAGCATCC 180
GGCACCCCGA GGAGCTGTCC CTGCTCCGGG CTCCTGAGAA GAAGGAGAAG AAGAAGAAAG 240
AGAAGGAGCC AGAGGAAGAG CTCTATGACT TGAGCAAGGT TGTCTTGGCT GGGGGCGTGG 300
CACCTGCACT GTTCCGGGGG ATGCCAGCTC ACTTCTCGGA CAGCGCCCAG ACTGAGGCCT 360

GCTACCACAT GCTGAGCCGG CCCCAGCCGC CACCCGACCC CCTCCTGCTC CAGCGTCTGC 420
CACGGCCCAG CTCCCTGTCA GACAAGACCC AGCTCCACAG CAGGTGGCTG GACTCGTCGC 480
GGTGTCTCAT GCAGCAGGGC ATCAAGGCCG GGGACGCACT CTGGCTGCGC TTCAAGTACT 540
ACAGCTTCTT CGATTTGGAT CCCAAGACAG ACCCCGTGCG GCTGACACAG CTGTATGAGC 600
AGGCCCGGTG GGACCTGCTG CTGGAGGAGA TTGACTGCAC CGAGGAGGAG ATGATGGTGT 660
TTGCCGCCCT GCAGTACCAC ATCAACAAGC TGTCCCAGAG CGGGGAGGTG GGGGAGCCGG 720
CTGGCACAGA CCCAGGGCTG GACGACCTGG ATGTGGCCCT GAGCAACCTG GAGGTGAAGC 780
TGGAGGGGTC GGCGCCCACA GATGTGCTGG ACAGCCTCAC CACCATCCCA GAGCTCAAGG 840
ACCATCTCCG AATCTTTCGG CCCCGGAAGC TGACCCTGAA GGGCTACCGC CAACACTGGG 900
TGGTGTTCAA GGAGACCACA CTGTCCTACT ACAAGAGCCA GGACGAGGCC CCTGGGGACC 960
CCATTCAGCA GCTCAACCTC AAGGGCTGTG AGGTGGTTCC CGATGTTAAC GTCTCCGGCC 1020
AGAAGTTCTG CATTAAACTC CTAGTGCCCT CCCCTGAGGC ATGAGTGAGA TCTACCTGCG 1080
GTGCCAGGAT GAGCAGCAGT ATGCCCGCTG GATGGCTGGC TGCCGCCTGG CCTCCAAAGG 1140
CCGCACCATG GCCGACAGCA GCTACACCAG CGAGGTGCAG GCCATCCTGG CYTTCCTCAG 1200
CCTGCAGCGC ACGGGCAGTG GGGGCCCGGG CAACCACCCC CACGGCCCTG ATGCCTCTGC 1260
CGAGGGCCTC AACCCCTACG GCCTCGTTGC CCCCCGTTTC CAGCGAAAGT TCAAGGCCAA 1320
GCAGCTCACC CCACGGATCC TGGAAGCCCA CCAGAATGTG GCCCAGTTGT CGCTGGCAGA 1380
GGCCCAGCTG CGCTTCATCC AGGCCTGGCA GTCCCTGCCC GACTTCGGCA TCTCCTATGT 1440
CATGGTCAGG TTCAAGGGCA GCAGGAAAGA CGAGATCCTG GGCATCGCCA ACAACCGACT 1500
GATCCGCATC GACTTGGCCG TGGGCGACGT GGTCAAGACC TGGCGTTTCA GCAACATGCG 1560
CCAGTGGAAT GTCAACTGGG ACATCCGGCA NGTGGCCATC GAGTTTGATG AACACATCAA 1620
TGTGGCCTTC AGCTGCGTGT CTGCCAGCTG CCGAATTGTA CACGAGTATA TCGGGGGCTA 1680
CATTTTCCTG TCGACGCGGG AGNGGGCCCG TGGGGAGGAG CTGGATGAAG ACCTCTTCCT 1740
GCAGCTCACC GGGGGCCATG AGGCCTTCTG AGGGCTGTCT GATTGCCCCT GCCCTGCTCA 1800
CCACCCTGTC ACAGCCACTC CCAAGCCCAC ACCCACAGGG GCTCACTGCC CCACACCCGC 1860
TCCAGGCAGG CACCCAGCTG GGCATTTCAC CTGCTGTCAC TGACTTTGTG CAGGCCAAGG 1920
ACCTGGCAGG GCCAGACGCT GTACCATCAC CCAGGCCAGG GATGGGGGTG GGGGTCCCTG 1980
AGCTCATGTG GTGCCCCCTT TCCTTGTCTG AGTGGCTGAG GCTGATACCC CTGACCTATC 2040
TGCAGTCCCC CAGCACACAA GGAAGACCAG ATGTAGCTAC AGGATGATGA AACATGGTTT 2100
CAAACGAGTT CTTTCTTGTT ACTTTTTAAA ATTTCTTTTT TATAAATTAA TATTTTATTG 2160

TTGGATCCTC CTCCTTTCTC TGGAGCTGTG CTTGGGGCTA CTCTGACACT CTGTCTCTTC 2220

ATCACCAGCC AAGGAAAGGG GCTTTCCTGA TAAAGACAAG AGTTGGTTAG AGAAAGGGAC 2280

ACCTAAGTCA GTCTAGGGTT GGAAGCTAGG AGAGAGGTGA GGGCAGAAGG GCACAGCTTT 2340

CAGGAACAAG GAATAGGGGC TGGGGTKGTK GTTCTCACGG GTAGGCGGTA CCTGCAGGGC 2400

CTCCTTGAAG TACTTGGGAA GGAGGAAGCC ATCAGTATTC CCTGGAGTCA GAATCACCCC 2460

ATTGGCAGAG CGGAAGAAGG GTATTCCATC TGCTGACAGA GCCAGAGATG TGACTCATGC 2520

CCTCCCCGAA GGCAAAGTCA GCTCCTGCTT TGTCCAGACT CACCTGCCAG AGCCAGGGGT 2580

C 2581

(2) INFORMATION FOR SEQ ID NO: 44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1764 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

GAATTCGGCA CGAGGATGAT ATTCCTACTA TTCCTCACCC CACTCTGGCT GCAAAAAGGA 60

AGTGCAGGGA AAATGAGTGG GGAGTTCCTG TATGCCAGTC TGTTTCAATG GAACTATTTT 120

TGGAGGAATA AAAAAGTATG CTAGATTATA TTGGTACGAT AGGCATTTTC TTACATTGCA 180

TATAGTCTGC TTTGGCTTTT ACCTGTTGAG GGGAAGAATG AGGAGAGGAT AAAAATCATT 240

GTATCCCCTA GAGAAGGAAT ATCAAAATCC ATTTAATAAA AAAACTCATA CTAAGAATAA 300

AATTGCATAG TGTTTTATTC TCCTTTGTTC ATAATTAAAC ACAAGATATT TTAAATTGTC 360

AAATCAGTTT CTTTATGAAA AAATATGACC TGTATGCCTT TATTCTCTCC TTTCCTTCTT 420

CCCACCCGTC GCTTCTTTTC TTCTCTTCCT TTTTTTCTTT CCTTGTCCTC TGACTAAATG 480

AAGAACAAAC ATTTGATAAA AGCCACTGCC AATTCATGAT AAAAATTCAC AGCAAAGTTG 540

GTACAGAAAA GAACTTTCTC TGCGTGTTAA AGGGTGCCTC TCCCATGCTC TCAGCAAATA 600

TTTAATGATG AAATCTTATT AATAATCACT GTAGAACCAA GAATTAAACT AGTATACCCA 660

CTGTCTTGGC TTGTAATCAA CAATATACAG GTGGTTCTAG CCAGTGCAAT AAGACAAGAG 720

AAACAAAAAT GTTATAAGGC CTGGAAAAGA TGAAACAAAC TGTTATTCAC AAAATACTGT 780

CTATACAGAA TGCTCAGTGT CTTTTTTTCT TTTCTTTTTT TTAAACTTTA GTGAGATACC 840

CTTCTGCCCT ATCTTAAAAT CACGTGGTGG GGGGTGGTGT CTGCACTTGA AACAGGACAC 900

```
TTGGTTCCTG GGTTTAGCAT TGACCTTGCC AGCTTGGTYT GGCAGCTGAG TTGTTGGACT     960
AGGAAGCGTC CYTGCAGGTT GTGKTCTGKT ACCTCTCTGT AAAGCCTGAA AGCATCCTAC    1020
SATTGCATTT GCTAGKTCTC AGTAGAGCTA TTTAACAAGA ATCTGGAAAC ATTTTYCCTG    1080
AGGGCTCTCT TTAGACAGCA GTAAAATGTA GCTGGAGACA TATTGAGTAA ATGGAAAAGA    1140
AAAATCTAAT GAGGCCAGGA ATTTTTTTAA TCTTCTATTC TCACAGAAGG CCTCAAGGAG    1200
AACACCATAA TTCATATTTT ACTCAKGTGG GTTAGGCATA AAGCCTCCCC CATAGATCCA    1260
ATAACCTGTA RGTGTYCTGG TTTTGAAATT GCACCTGCTT ACATKGCTGG ATCNTAGCAC    1320
TAAWTCACAC RGCAACGGCT TCTGGTTCAA TKGTTCATTA CTTGGGAATG TCAGATTGCC    1380
AGAGAGCAGC CTGATGTTTA CATCCAATCG GCAATGCCTT AGGAAATCAG TTTTAATTAC    1440
AATCTCACGT AGCAGCACTG CACTCAACCT TCAGAGAGGC TGGGATTTGT GTTGAACCTA    1500
CATCTTATAG CTGTGCAGAA AATGCCTGTC CGACTGGGTC ATGCAAAATG GACAGCAAAG    1560
TCAGCAGAAC CTTAGAAAAG ATGACACAGC AAGTGGAACA CAGCTGGATC ATCCCCCGTC    1620
CTGTCAAGCG TGCAGTGCTC TCTGGCCCCT TTTTAAAACA AGGGAACCCA GTTGGCGTTT    1680
GCCTTTCAGC TTCCCCATTC TGATATAAAA ATCTGTGACC CAGCAGCTTT AACCATAAAA    1740
AAAAAAAAAA AAAAAAAAAC TCGA                                           1764
```

(2) INFORMATION FOR SEQ ID NO: 45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 796 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

```
ACCTTCTTCC ATGTTTAGTC CCTTGGGCTC TGCTACCCTC CTGCTGGAGG TGAGAGCATC      60
CTGTGTGCAA CCAGAGATGC CCTCTGGCTT TCAGACCTGC CTGCTTTTCA CCCTCAGCCC     120
TTTCTCACTC AGCAAAATTG TGGGGGTCCC TAGTCAGCAG CTCCCTGGGC AGCTCTCTGA     180
GCAAGGTGGT CTCTGTGGTC ATGAAGGAGA GCCGGCTAGG ACAGTGCCGG AAACTCAGCT     240
GCCTCTCCCC TTCAACTCAG CTGGCCCCCC GCACCTGAAG TGCACAGGAG CCGGGAAGAG     300
AGTCTGGAGC CCACCCCGGA GGGCAGCACA GGAGGTGTCT YTGCAGCTGG TGTCCTGCMA     360
CCCYTGCAGG CAGMACACGT CCCGGGCATT YTCYTTAGCC ACAGACAGAA CAGCCAGTGC     420
CAGAGTCTGC TGTCGYTTCC CCTTTAAGCA CACTCATTCA CCACACCCGA GGAGGCCAGA     480
GGTGCAGGGA GCATGGGCTG TCGTTCCCCT TTAAGCACAC TCATTCACCA CACCCGAGGA     540
```

GGCCAGAAGT GCAGGGAGCA TGGGCTGGGT GCACCTCCGC AGGAGAGAAG GCTGAGCCAC 600

CGCCGTCCCG GGAGCCCGGC TCCCAGGCCT CTCGTTTTCC CCTACCTCCC TAAGACTTTT 660

CTGTCACTCT CTGGCCATTG AAAGGCTTCT GTTCCTTAAA GTGCTGTTAC ACTCTCCTTT 720

CCCAGGATGC AGCAAGCCAA AACAGTACCA CTGCACGTCA GCCTGGGTGA CAGAGTGAGA 780

CCCTATCTTA AAAAAA 796

(2) INFORMATION FOR SEQ ID NO: 46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1705 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

TGGCCATGGA AGCGCTAGAA GGTTTAGATT TTGAAACAGC AAAGAAGGAT TTCCTTGGAT 60

CTGGAGACCC CAAAGAAACA AAGATGCTAA TCACCAAACA GGCTGACTGG GCCAGAAATA 120

TCAAGGAGCC CAAAGCCGCC GTGGAGATGT ACATCTCAGC AGGAGAGCAC GTCAAGGCCA 180

TCGAGATCTG TGGTGACCAT GGCTGGGTTG ACATGTTGAT CGACATCGCC CGCAAACTGG 240

ACAAGGCTGA GCGCGAGCCC CTGCTGCTGT GCGCTACCTA CCTCAAGAAG CTGGACAGCC 300

CTGGCTATGC TGCTGAGACC TACCTGAAGA TGGGTGACCT CAAGTCCCTG GTGCAGCTGC 360

AGTGGAGACC CAGCGCTGGG ATGAGGCCTT TGCTTTGGGT GAGAAGCATC CTGAGTTTAA 420

GGATGACATC TACATGCCGT ATGCTCAGTG GCTAGCAGAG AACGATCGCT TTGAGGAAGC 480

CCAGAAAGCG TTCCACAAGG CTGGGCGACA GAGAGAAGCG GTCCAGGTGC TGGAGCAGCT 540

CACAAACAAT GCCGTGGCGG AGAGCAGGTT TAATGATGCT GCCTATTATT ACTGGATGCT 600

GTCCATGCAG TGCCTCGATA TAGCTCAAGA TCCTGCCCAG AAGGACACAA TGCTTGGCAA 660

GTTCTACCAC TTCCAGCGTT TGGCAGAGCT GTACCATGGT TACCATGCCA TCCATCGCCA 720

CACGGAAGAT CCGTTCAGTG TCCATCGTCC TGAAACTCTT TTCAACATCT CCAGGTTCCT 780

GCTGCACAGC CTGCCCAAGG ACACCCCCTC GGGCATCTCT AAAGTGAAAA TACTCTTCAC 840

CTTGGCCAAG CAGAGCAAGG CCCTCGGTGC CTACAGGCTG GCCCGGCACG CCTATGACAA 900

GCTGCGTGGC CTGTACATCC CTGCCAGATT CCAAAAGTCC ATTGAGCTGG GTACCCTGAC 960

CATCCGCGCC AAGCCCTTCC ACGACAGTGA GGAGTTGGTG CCCTTGTGCT ACCGCTGCTC 1020

CACCAACAAC CCGCTGCTCA ACAACCTGGG CAACGTCTGC ATCAACTGCC GCCAGCCCTT 1080

CATCTTCTCC GCCTCTTCCT ACGACGTGCT ACACCTGGTT GAGTTCTACC TGGAGGAAGG 1140

GATCACTGAT GAAGAAGCCA TCTCCCTCAT CGACCTGGAG GTGCTGAGAC CCAAGCGGGA 1200

TGACAGACAG CTAGAGATTT GCAAACAACA GCTCCCAGAT TCTTGCGGCT AGTGGGAGAC 1260

CAAGGGACTC CATCGGAGAT NAGGACCCGT TCACAGCTAA GCTRAGCTTT GAGCAAGGTG 1320

GCTCARAGTT CGTGCCAGTG GTGGTGAGCC GGCTGGTGCT GCGCTCCATG AGCCGCCGGG 1380

ATGTCCTCAT CAAGCGATGG CCCCCACCCC TGAGGTGGCA ATACTTCCGC TCACTGCTGC 1440

CTGACGCCTC CATTACCATG TGCCCCTCCT GCTTCCAGAT GTTCCATTCT GAGGACTATG 1500

AGTTGCTGGT GCTTCAGCAT GGCTGCTGCC CCTACTGCCG CAGGTGCAAG GATGACCCTG 1560

GCCCATGACC AGCATCCTGG GGACGGCCTG CACCCTCTGC CCGCCTTGGG GTCTGCTGGG 1620

CTGTGAAGGA GAATAAAGAG TTAAACTGTC AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1680

AAAAAAAAAA AAAAAAAAAA AAANA 1705

(2) INFORMATION FOR SEQ ID NO: 47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 981 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

TCGGCAGCAC AGAGCTCTGG AGATGAAGAC CCTGTTCCTG GGTGTCACGC TCGGCCTGGC 60

GCTGCCCTGT CCTTCACCCT GGRGGAGGAG GATATCACAG GGACCTGGTA CGTGAAGGCC 120

ATGGTGGTCG ATAAGACTTT CCGGAGACAG GAGGCCCAGA AGGTGTCCCC AGTGAAGGTG 180

ACAGCCCTGG GCGGTGGGAA GTTGGAAGCC ACGTTCACCT TCATGAGGGA GGATCGGTGC 240

ATCCAGAAGA AAATCCTGRT GCGGAAGACG GAGGAGCCTG GCAAATACAG CGCCTGTGAG 300

CCCCTCCCCC AYTCCCACCC CCACCYTCCC CCACCGCCAA CCCCAGTGCA CCAGCCTCCA 360

CAGGTAGAGA GTGCCCAGGC TGCCCTTTTG CCAGGGCCCC AGCTCTGCCC ACCTCCAAGG 420

AGGGGCTGGC CTCTCCTTCC TGGGGGGCTG GTGGCCCTGA CATCAGACAC CGGGTGTGAC 480

AGGCTTGTCC GCAGTCGAGA TGGACCAGAT CACGCCTGCC CTCTGGGAGG CCCTAGCCAT 540

TGACACATTG AGGAAGCTGA GGATTGGGAC AAGGAGGCCA AGGATTAGAT GGGGGCAGGA 600

AGCTCATGTA CCTGCAGGAG CTGCCCAGGA GGGACCAYTA CATCTTTTAC TGCAAAGACC 660

AGCACCATGG GGGCSTGCTC CACATGGGAA AGCTTGTGGG TAGGAATTCT GATACCAACC 720

GGGAGGCCCT GGAAGAATTT AAGAAATTGG TGCAGCGCAA GGGACTCTCG GAGGAGGACA 780

TTTTCACGCC CCTGCAGACG GGAAGCTGCR TTCCCGAACA CTAGGCAGCC CCCGGGTCTG 840

CACCTCCAGA GCCCACCCTA CCACCAGACA CAGAGCCCGG ACCACCTGGA CCTACCCTCC 900

AGCCATGACC CTTCCCTGCT CCCACCCACC TGACTCCAAA TAAAGTCCTT CTCCCCCAAA 960

AAAAAAAAAA AAAAAACTCG A 981

(2) INFORMATION FOR SEQ ID NO: 48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 146 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

```
Met His Tyr Gln Met Ser Val Thr Leu Lys Tyr Glu Ile Lys Lys Leu
  1               5                  10                  15

Ile Tyr Val His Leu Val Ile Trp Leu Leu Leu Val Ala Lys Met Ser
             20                  25                  30

Val Gly His Leu Arg Leu Leu Ser His Asp Gln Val Ala Met Pro Tyr
         35                  40                  45

Gln Trp Glu Tyr Pro Tyr Leu Leu Ser Ile Leu Pro Ser Leu Leu Gly
     50                  55                  60

Leu Leu Ser Phe Pro Arg Asn Asn Ile Ser Tyr Leu Val Leu Ser Met
 65                  70                  75                  80

Ile Ser Met Gly Leu Phe Ser Ile Ala Pro Leu Ile Tyr Gly Ser Met
                 85                  90                  95

Glu Met Phe Pro Ala Ala Gln Pro Ser Thr Ala Met Ala Arg Pro Thr
            100                 105                 110

Val Ser Ser Leu Val Phe Leu Pro Phe Pro Ser Cys Thr Trp Cys Trp
        115                 120                 125

Cys Trp Gln Cys Lys Cys Met Pro Gly Ser Cys Thr Thr Ala Arg Ser
    130                 135                 140

Ser Xaa
145
```

(2) INFORMATION FOR SEQ ID NO: 49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 312 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

Met Asn Ser Val Val Ser Leu Leu Leu Ile Leu Glu Pro Asp Lys Gln
1 5 10 15

Glu Ala Leu Ile Glu Ser Leu Cys Glu Lys Leu Val Lys Phe Arg Glu
20 25 30

Gly Glu Arg Pro Ser Leu Arg Leu Gln Leu Leu Ser Asn Leu Phe His
35 40 45

Gly Met Asp Lys Asn Thr Pro Val Arg Tyr Thr Val Tyr Cys Ser Leu
50 55 60

Ile Lys Val Ala Ala Ser Cys Gly Ala Ile Gln Tyr Ile Pro Thr Glu
65 70 75 80

Leu Asp Gln Val Arg Lys Trp Ile Ser Asp Trp Asn Leu Thr Thr Glu
85 90 95

Lys Lys His Thr Leu Leu Arg Leu Leu Tyr Glu Ala Leu Val Asp Cys
100 105 110

Lys Lys Ser Asp Ala Ala Ser Lys Val Met Val Glu Leu Leu Gly Ser
115 120 125

Tyr Thr Glu Asp Asn Ala Ser Gln Ala Arg Val Asp Ala His Arg Cys
130 135 140

Ile Val Arg Ala Leu Lys Asp Pro Asn Ala Phe Leu Phe Asp His Leu
145 150 155 160

Leu Thr Leu Lys Pro Val Lys Phe Leu Glu Gly Glu Leu Ile His Asp
165 170 175

Leu Leu Thr Ile Phe Val Ser Ala Lys Leu Ala Ser Tyr Val Lys Phe
180 185 190

Tyr Gln Asn Asn Lys Asp Phe Ile Asp Ser Leu Gly Leu Leu His Glu
195 200 205

Gln Asn Met Ala Lys Met Arg Leu Leu Thr Phe Met Gly Met Ala Val
210 215 220

Glu Asn Lys Glu Ile Ser Phe Asp Thr Met Gln Gln Glu Leu Gln Ile
225 230 235 240

Gly Ala Asp Asp Val Glu Ala Phe Val Ile Asp Ala Val Arg Thr Lys
245 250 255

Met Val Tyr Cys Lys Ile Asp Gln Thr Gln Arg Lys Val Val Val Ser
260 265 270

His Ser Thr His Arg Thr Phe Gly Lys Gln Gln Trp Gln Gln Leu Tyr
275 280 285

Asp Thr Leu Asn Ala Trp Lys Gln Asn Leu Asn Lys Val Lys Asn Ser
290 295 300

Leu Leu Ser Leu Ser Asp Thr Xaa
305 310

(2) INFORMATION FOR SEQ ID NO: 50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 47 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

```
Gly Gly Cys Pro Arg Arg Arg Leu Val Leu Tyr Cys Leu Phe Gly Ser
 1               5                  10                  15

Ala Gly Gly Gly Arg Ile His Ser Glu Ala Trp Phe Pro Lys Ala Trp
            20                  25                  30

Pro Glu Ala Glu Lys Trp Leu Phe Ala Glu Leu Leu Arg Gly Xaa
        35                  40                  45
```

(2) INFORMATION FOR SEQ ID NO: 51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 467 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

```
Met Leu Ser Arg Pro Gln Pro Pro Pro Asp Pro Leu Leu Leu Gln Arg
 1               5                  10                  15

Leu Pro Arg Pro Ser Ser Leu Ser Asp Lys Thr Gln Leu His Ser Arg
            20                  25                  30

Trp Leu Asp Ser Ser Arg Cys Leu Met Gln Gln Gly Ile Lys Ala Gly
        35                  40                  45

Asp Ala Leu Trp Leu Arg Phe Lys Tyr Tyr Ser Phe Phe Asp Leu Asp
    50                  55                  60

Pro Lys Thr Asp Pro Val Arg Leu Thr Gln Leu Tyr Glu Gln Ala Arg
65                  70                  75                  80

Trp Asp Leu Leu Leu Glu Glu Ile Asp Cys Thr Glu Glu Glu Met Met
                85                  90                  95

Val Phe Ala Ala Leu Gln Tyr His Ile Asn Lys Leu Ser Gln Ser Gly
            100                 105                 110

Glu Val Gly Glu Pro Ala Gly Thr Asp Pro Gly Leu Asp Asp Leu Asp
        115                 120                 125

Val Ala Leu Ser Asn Leu Glu Val Lys Leu Glu Gly Ser Ala Pro Thr
    130                 135                 140

Asp Val Leu Asp Ser Leu Thr Thr Ile Pro Glu Leu Lys Asp His Leu
145                 150                 155                 160

Arg Ile Phe Arg Pro Arg Lys Leu Thr Leu Lys Gly Tyr Arg Gln His
```

```
                165                     170                     175

Trp Val Val Phe Lys Glu Thr Thr Leu Ser Tyr Tyr Lys Ser Gln Asp
            180                     185                     190

Glu Ala Pro Gly Asp Pro Ile Gln Gln Leu Asn Leu Lys Gly Cys Glu
        195                     200                     205

Val Val Pro Asp Val Asn Val Ser Gly Gln Lys Phe Cys Ile Lys Leu
    210                     215                     220

Leu Val Pro Ser Pro Glu Gly Met Ser Glu Ile Tyr Leu Arg Cys Gln
225                     230                     235                     240

Asp Glu Gln Gln Tyr Ala Arg Trp Met Ala Gly Cys Arg Leu Ala Ser
                245                     250                     255

Lys Gly Arg Thr Met Ala Asp Ser Ser Tyr Thr Ser Glu Val Gln Ala
            260                     265                     270

Ile Leu Ala Phe Leu Ser Leu Gln Arg Thr Gly Ser Gly Gly Pro Gly
        275                     280                     285

Asn His Pro His Gly Pro Asp Ala Ser Ala Glu Gly Leu Asn Pro Tyr
    290                     295                     300

Gly Leu Val Ala Pro Arg Phe Gln Arg Lys Phe Lys Ala Lys Gln Leu
305                     310                     315                     320

Thr Pro Arg Ile Leu Glu Ala His Gln Asn Val Ala Gln Leu Ser Leu
                325                     330                     335

Ala Glu Ala Gln Leu Arg Phe Ile Gln Ala Trp Gln Ser Leu Pro Asp
            340                     345                     350

Phe Gly Ile Ser Tyr Val Met Val Arg Phe Lys Gly Ser Arg Lys Asp
        355                     360                     365

Glu Ile Leu Gly Ile Ala Asn Asn Arg Leu Ile Arg Ile Asp Leu Ala
    370                     375                     380

Val Gly Asp Val Val Lys Thr Trp Arg Phe Ser Asn Met Arg Gln Trp
385                     390                     395                     400

Asn Val Asn Trp Asp Ile Arg Gln Val Ala Ile Glu Phe Asp Glu His
                405                     410                     415

Ile Asn Val Ala Phe Ser Cys Val Ser Ala Ser Cys Arg Ile Val His
            420                     425                     430

Glu Tyr Ile Gly Gly Tyr Ile Phe Leu Ser Thr Arg Glu Arg Ala Arg
        435                     440                     445

Gly Glu Glu Leu Asp Glu Asp Leu Phe Leu Gln Leu Thr Gly Gly His
    450                     455                     460

Glu Ala Phe
465
```

(2) INFORMATION FOR SEQ ID NO: 52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 83 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

```
Met Arg Pro Gly Arg Gly Ala Gly Thr Pro Gly Arg Pro Gly Arg Gly
 1               5                  10                  15

Arg Gly Leu Ala Ala Thr Cys Ser Leu Ser Ser Pro Ser His Leu Leu
            20                  25                  30

Pro Thr Leu Leu His Thr Phe Ser Phe Ser Leu Pro Pro Pro Ser Pro
        35                  40                  45

Ala Ala Pro Arg Gln Pro Ser Pro Pro Ala Leu Leu Leu Pro Gly Pro
    50                  55                  60

Gln Lys Pro Arg Pro Gly Asp Pro Thr Tyr Thr Gly Ala Leu Thr Asp
65                  70                  75                  80

Trp Ser Xaa
```

(2) INFORMATION FOR SEQ ID NO: 53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 63 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

```
Met Phe Leu Val Phe Phe Leu Ser Phe Phe Ser His Ser Ile Ser Ala
 1               5                  10                  15

Leu Thr Leu Val Cys Ser Gln Gly Gly Lys Ala Asp Met Asn Leu Leu
            20                  25                  30

Ser Trp Asp Phe Arg Pro His Trp Leu Glu Gly Ile Arg Phe Leu Leu
        35                  40                  45

Gly Trp Gly Gln Ala Leu Met Ala Gly Leu Phe Pro Trp Leu Xaa
    50                  55                  60
```

(2) INFORMATION FOR SEQ ID NO: 54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 124 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

```
Met Arg Gly Ser Trp His Arg Ser Pro Leu Pro Ala Val Val Leu Pro
```

```
1               5                   10                  15
Ser Val Leu Gln Thr Ala Leu Ser Pro Leu Ala Leu Cys Gln Ala Trp
                20                  25                  30
Arg Arg Ala Val Pro His Gly Val Pro Ser Gln Arg Leu Arg Asn Gln
            35                  40                  45
Glu Ala Ser Leu Val Pro Lys Gly Val Pro Arg Ala Trp Tyr Pro Gly
    50                  55                  60
Pro Leu Gln Asn Gly Leu Trp Thr His Leu Glu Lys Gly Glu Leu Leu
65                  70                  75                  80
Gly Leu Lys Pro Thr Pro Gly Gly Leu Leu Leu Leu Arg Ser Phe Trp
                85                  90                  95
Asp Pro His Pro Ser Arg Pro Phe Leu Cys Thr Leu Leu Pro Pro Pro
            100                 105                 110
Leu Xaa Ile Phe Pro Pro Leu Arg Cys Ser Ala Xaa
        115                 120
```

(2) INFORMATION FOR SEQ ID NO: 55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 180 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

```
Met Thr Ser Ala Gly Pro Val Xaa Leu Phe Leu Leu Val Ser Ile Ser
1               5                   10                  15
Thr Ser Val Ile Leu Met Gln His Leu Leu Xaa Ala Ser Tyr Cys Asp
            20                  25                  30
Leu Leu His Lys Ala Ala Ala His Leu Gly Cys Trp Gln Lys Val Asp
        35                  40                  45
Pro Ala Leu Cys Ser Asn Val Leu Gln His Pro Trp Thr Glu Glu Cys
    50                  55                  60
Met Trp Pro Gln Gly Val Leu Val Lys His Ser Lys Asn Val Tyr Lys
65                  70                  75                  80
Ala Val Gly Xaa Xaa Xaa Val Ala Ile Pro Ser Asp Val Ser His Phe
                85                  90                  95
Arg Phe Xaa Phe Phe Phe Ser Lys Pro Leu Arg Ile Leu Asn Ile Leu
            100                 105                 110
Leu Leu Leu Glu Gly Ala Val Ile Val Tyr Gln Leu Tyr Ser Leu Met
        115                 120                 125
Ser Ser Glu Lys Trp His Gln Thr Ile Ser Leu Ala Leu Ile Leu Phe
    130                 135                 140
```

Ser Asn Tyr Tyr Ala Phe Phe Lys Leu Leu Arg Asp Arg Leu Val Leu
145 150 155 160

Gly Lys Ala Tyr Ser Tyr Ser Ala Ser Pro Gln Arg Asp Leu Asp His
165 170 175

Arg Phe Ser Xaa
180

(2) INFORMATION FOR SEQ ID NO: 56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 287 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

Met Pro Leu Phe Lys Leu Tyr Met Val Met Ser Ala Cys Phe Leu Ala
1 5 10 15

Ala Gly Ile Phe Trp Val Ser Ile Leu Cys Arg Asn Thr Tyr Ser Val
20 25 30

Phe Lys Ile His Trp Leu Met Ala Ala Leu Ala Phe Thr Lys Ser Ile
35 40 45

Ser Leu Leu Phe His Ser Ile Asn Tyr Tyr Phe Ile Asn Ser Gln Gly
50 55 60

Pro Pro His Arg Arg Pro Cys Arg His Val Leu His Arg Thr Pro Ala
65 70 75 80

Glu Gly Arg Pro Pro Leu His His His Arg Pro Asp Trp Leu Arg Leu
85 90 95

Gly Phe Ile Lys Tyr Val Leu Ser Asp Lys Glu Lys Lys Val Phe Gly
100 105 110

Ile Val Ile Pro Met Gln Val Leu Ala Asn Val Ala Tyr Ile Ile Ile
115 120 125

Glu Ser Arg Glu Glu Gly Ala Thr Asn Tyr Val Leu Trp Lys Glu Ile
130 135 140

Leu Phe Leu Val Asp Leu Ile Cys Cys Gly Ala Ile Leu Phe Pro Val
145 150 155 160

Val Trp Ser Ile Arg His Leu Gln Asp Ala Ser Gly Thr Asp Gly Lys
165 170 175

Val Ala Val Asn Leu Ala Lys Leu Lys Leu Phe Arg His Tyr Tyr Val
180 185 190

Met Val Ile Cys Tyr Val Tyr Phe Thr Arg Ile Ile Ala Ile Leu Leu
195 200 205

Gln Val Ala Val Pro Phe Gln Trp Gln Trp Leu Tyr Xaa Leu Leu Val
210 215 220

```
Glu Gly Ser Thr Leu Ala Phe Phe Val Leu Thr Gly Tyr Lys Phe Gln
225                 230                 235                 240

Pro Thr Gly Asn Asn Pro Tyr Leu Gln Leu Pro Gln Glu Asp Glu Glu
                245                 250                 255

Asp Val Gln Met Glu Gln Val Met Thr Asp Ser Gly Phe Arg Glu Gly
            260                 265                 270

Leu Ser Lys Val Asn Lys Thr Ala Ser Gly Arg Glu Leu Leu Xaa
        275                 280                 285
```

(2) INFORMATION FOR SEQ ID NO: 57:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

```
Met Pro Met Val Phe Leu Leu Leu Phe Asn Leu Met Ser Trp Leu Ile
  1               5                  10                  15

Arg Asn Ala Arg Val Ile Leu Arg Ser Leu Asn Leu Lys Arg Asp Gln
            20                  25                  30

Val Xaa
```

(2) INFORMATION FOR SEQ ID NO: 58:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

```
Met Lys Ile Val Val Leu Leu Pro Leu Phe Leu Leu Ala Thr Phe Pro
  1               5                  10                  15

Arg Lys Leu Gln Thr Cys Leu Xaa
            20
```

(2) INFORMATION FOR SEQ ID NO: 59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 47 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

```
Met Ser Gly Gly Glu Gly Ala Ala Leu Pro Ile Leu Leu Leu Leu Leu
  1               5                  10                  15
```

```
Ala Leu Arg Gly Thr Phe His Gly Ala Arg Pro Gly Gly Gly Ala Ser
            20                  25                  30

Gly Ile Trp Cys Leu Leu Leu Pro Glu Gln Glu Pro Pro Val Xaa
        35                  40                  45
```

(2) INFORMATION FOR SEQ ID NO: 60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 114 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

```
Met Ala Arg Gly Ser Leu Arg Arg Leu Leu Arg Leu Leu Val Leu Gly
  1               5                  10                  15

Leu Trp Leu Ala Leu Leu Arg Ser Val Ala Gly Glu Gln Ala Pro Gly
            20                  25                  30

Thr Ala Pro Cys Ser Arg Gly Ser Ser Trp Ser Ala Asp Leu Asp Lys
        35                  40                  45

Cys Met Asp Cys Ala Ser Cys Arg Ala Arg Pro His Ser Asp Phe Cys
    50                  55                  60

Leu Gly Cys Ala Ala Ala Pro Pro Ala Pro Phe Arg Leu Leu Trp Pro
 65                  70                  75                  80

Ile Leu Gly Gly Ala Leu Ser Leu Thr Phe Val Leu Gly Leu Leu Ser
                 85                  90                  95

Gly Phe Leu Val Trp Arg Arg Cys Arg Arg Glu Arg Ser Ser Pro Pro
            100                 105                 110

Pro Xaa
```

(2) INFORMATION FOR SEQ ID NO: 61:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

```
Met Val Cys Ile Leu Val Leu Thr Leu Val Ser Tyr Ser Ser Leu Val
  1               5                  10                  15

Asn Ser Pro Leu Pro Phe Val His Leu Xaa Val Gly Ile Ser Ala Xaa
            20                  25                  30
```

(2) INFORMATION FOR SEQ ID NO: 62:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 81 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

```
Met Thr Gly Gly Phe Leu Ser Cys Ile Leu Gly Leu Val Leu Pro Leu
 1               5                  10                  15

Ala Tyr Xaa Ser Ser Leu Thr Trp Cys Trp Trp Arg Trp Gly Leu Pro
            20                  25                  30

Xaa Pro Ala Gly Pro Pro Arg Cys Thr Pro Gly Cys Asn Ala Ser Gly
        35                  40                  45

Ala Gly Arg Gly Pro Ser Pro Gly Pro Pro Gly Gly Glu Leu His Thr
    50                  55                  60

Pro Ala Ser Arg Asp Pro Gly Pro Gly Ala Glu Trp Arg Gly Thr Ser
65                  70                  75                  80

Xaa
```

(2) INFORMATION FOR SEQ ID NO: 63:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 104 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

```
Met Ala Ala Pro Val Asp Leu Glu Leu Lys Lys Ala Phe Thr Glu Leu
 1               5                  10                  15

Gln Ala Lys Val Ile Asp Thr Gln Gln Lys Val Lys Leu Ala Asp Ile
            20                  25                  30

Gln Ile Glu Gln Leu Asn Arg Thr Lys Lys His Ala His Leu Thr Asp
        35                  40                  45

Thr Glu Ile Met Thr Leu Val Asp Glu Thr Asn Met Tyr Glu Gly Val
    50                  55                  60

Gly Arg Met Phe Ile Leu Gln Ser Lys Glu Ala Ile His Ser Gln Leu
65                  70                  75                  80

Leu Glu Lys Gln Lys Ile Ala Glu Glu Lys Ile Lys Glu Leu Glu Gln
                85                  90                  95

Lys Lys Ser Tyr Leu Glu Arg Arg
            100
```

(2) INFORMATION FOR SEQ ID NO: 64:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 146 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

```
Met Pro Ser Gly Phe Gln Thr Cys Leu Leu Phe Thr Leu Ser Pro Phe
 1               5                  10                  15

Ser Leu Ser Lys Ile Val Gly Val Pro Ser Gln Gln Leu Pro Gly Gln
            20                  25                  30

Leu Ser Glu Gln Gly Gly Leu Cys Gly His Glu Gly Glu Pro Ala Arg
        35                  40                  45

Thr Val Pro Glu Thr Gln Leu Pro Leu Pro Phe Asn Ser Ala Gly Pro
    50                  55                  60

Pro His Leu Lys Cys Thr Gly Ala Gly Lys Arg Val Trp Ser Pro Pro
65                  70                  75                  80

Arg Arg Ala Ala Gln Glu Val Ser Leu Gln Leu Val Ser Cys His Pro
                85                  90                  95

Cys Arg Gln His Thr Ser Arg Ala Phe Ser Leu Ala Thr Asp Arg Thr
            100                 105                 110

Ala Ser Ala Arg Val Cys Cys Arg Ser Pro Leu Ser Thr Leu Ile His
        115                 120                 125

His Thr Arg Gly Gly Gln Arg Cys Arg Glu His Gly Leu Ser Leu Pro
    130                 135                 140

Leu Xaa
145
```

(2) INFORMATION FOR SEQ ID NO: 65:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

```
Met Ala Ile Leu Met Leu Leu Ala Gly Ser Pro Cys Thr Leu Ser Phe
 1               5                  10                  15

Ser Thr Asp Thr Gly Ser Ser Ala Pro Gly Pro Lys Ile Pro Xaa
            20                  25                  30
```

(2) INFORMATION FOR SEQ ID NO: 66:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 260 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

Met Asp Pro Gln Gly Gln Thr Leu Leu Leu Phe Leu Phe Val Asp Phe
1 5 10 15

His Ser Ala Phe Pro Val Gln Gln Met Glu Ile Trp Gly Val Tyr Thr
20 25 30

Leu Leu Thr Thr His Leu Asn Ala Ile Leu Val Glu Ser His Ser Val
35 40 45

Val Gln Gly Ser Ile Gln Phe Thr Val Asp Lys Val Leu Glu Gln His
50 55 60

His Gln Ala Ala Lys Ala Gln Gln Lys Leu Gln Ala Ser Leu Ser Val
65 70 75 80

Ala Val Asn Ser Ile Met Ser Ile Leu Thr Gly Ser Thr Arg Ser Ser
85 90 95

Phe Arg Lys Met Cys Leu Gln Thr Leu Gln Ala Ala Asp Thr Gln Glu
100 105 110

Phe Arg Thr Lys Leu His Lys Val Phe Arg Glu Ile Thr Gln His Gln
115 120 125

Phe Leu His His Cys Ser Cys Glu Val Lys Gln Leu Thr Leu Glu Lys
130 135 140

Lys Asp Ser Ala Gln Gly Thr Glu Asp Ala Pro Asp Asn Ser Ser Leu
145 150 155 160

Glu Leu Leu Ala Asp Thr Ser Gly Gln Ala Glu Asn Lys Arg Leu Lys
165 170 175

Arg Gly Ser Pro Arg Ile Glu Glu Met Arg Ala Leu Arg Ser Ala Arg
180 185 190

Ala Pro Ser Pro Ser Glu Ala Ala Pro Arg Arg Pro Glu Ala Thr Ala
195 200 205

Ala Pro Leu Thr Pro Arg Gly Arg Glu His Arg Glu Ala His Gly Arg
210 215 220

Ala Leu Ala Pro Gly Arg Ala Ser Leu Gly Ser Arg Leu Glu Asp Val
225 230 235 240

Leu Trp Leu Gln Glu Val Ser Asn Leu Ser Glu Trp Leu Ser Pro Ser
245 250 255

Pro Gly Pro Xaa
260

(2) INFORMATION FOR SEQ ID NO: 67:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

```
Met Ala Ala Ala Cys Gly Pro Gly Ala Ala Gly Thr Ala Cys Ser Ser
 1               5                  10                  15

Ala Cys Ile Cys Phe Cys Xaa
            20
```

(2) INFORMATION FOR SEQ ID NO: 68:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

```
Met His Ala Leu Ile Leu Gln Phe Ile Phe Ser Leu Cys Met Tyr Ile
 1               5                  10                  15

Ser Leu Phe Ser Ala Ala Arg Phe Leu Phe Xaa
            20                  25
```

(2) INFORMATION FOR SEQ ID NO: 69:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

```
Leu Leu Leu Leu Cys Phe Cys Cys His Pro Thr His Leu Gln Gly Xaa
 1               5                  10                  15

Trp Ala Leu Asp Leu Gly Leu Phe Pro Phe Asn Cys Xaa
            20                  25
```

(2) INFORMATION FOR SEQ ID NO: 70:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 216 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

```
Met Tyr Leu Ser Ile Ile Phe Leu Ala Phe Val Ser Ile Asp Arg Cys
 1               5                  10                  15

Leu Gln Leu Thr His Ser Cys Lys Ile Tyr Arg Ile Gln Glu Pro Gly
            20                  25                  30
```

Phe Ala Lys Met Ile Ser Thr Val Val Trp Leu Met Val Leu Leu Ile
35 40 45

Met Val Pro Asn Met Met Ile Pro Ile Lys Asp Ile Lys Glu Lys Ser
50 55 60

Asn Val Gly Cys Met Glu Phe Lys Lys Glu Phe Gly Arg Asn Trp His
65 70 75 80

Leu Leu Thr Asn Phe Ile Cys Val Ala Ile Phe Leu Asn Phe Ser Ala
85 90 95

Ile Ile Leu Ile Ser Asn Cys Leu Val Ile Arg Gln Leu Tyr Arg Asn
100 105 110

Lys Asp Asn Glu Asn Tyr Pro Asn Val Lys Lys Ala Leu Ile Asn Ile
115 120 125

Leu Leu Val Thr Thr Gly Tyr Ile Ile Cys Phe Val Pro Tyr His Ile
130 135 140

Val Arg Ile Pro Tyr Thr Leu Ser Gln Thr Glu Val Ile Thr Asp Cys
145 150 155 160

Ser Thr Arg Ile Ser Leu Phe Lys Ala Lys Glu Ala Thr Leu Leu Leu
165 170 175

Ala Val Ser Asn Leu Cys Phe Asp Pro Ile Leu Tyr Tyr His Leu Ser
180 185 190

Lys Ala Phe Arg Ser Lys Val Thr Glu Thr Phe Ala Ser Pro Lys Glu
195 200 205

Thr Lys Val Arg Lys Lys Asn Xaa
210 215

(2) INFORMATION FOR SEQ ID NO: 71:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 407 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

Met His Pro Ala Val Phe Leu Ser Leu Pro Asp Leu Arg Cys Ser Leu
1 5 10 15

Leu Leu Leu Val Thr Trp Val Phe Thr Pro Val Thr Thr Glu Ile Thr
20 25 30

Ser Leu Asp Thr Glu Asn Ile Asp Glu Ile Leu Asn Asn Ala Asp Val
35 40 45

Ala Leu Val Asn Phe Tyr Ala Asp Trp Cys Arg Phe Ser Gln Met Leu
50 55 60

His Pro Ile Phe Glu Glu Ala Ser Asp Val Ile Lys Glu Glu Phe Pro
65 70 75 80

```
Arg Asp Arg Asp Glu Leu Xaa
                405
```

(2) INFORMATION FOR SEQ ID NO: 72:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

```
Tyr Leu Ile Ser Tyr Leu Cys Phe Xaa
 1               5
```

(2) INFORMATION FOR SEQ ID NO: 73:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

```
Met Pro Leu Lys Ala Val Thr Trp Pro Thr Leu Asn Ser Lys Leu Val
 1               5                  10                  15

Ala Ala Val Val Asn Leu Lys Ala Ser Gln Met Pro Ala Ser Ser Arg
            20                  25                  30

Val Xaa
```

(2) INFORMATION FOR SEQ ID NO: 74:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 57 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

```
Gln Ser Pro Arg Ser Ser Ala Leu Gly Ala Gly Gln Lys Leu Ala Val
 1               5                  10                  15

Cys Ser Pro Asp Ile Leu Cys Cys Pro Thr Asp Thr Leu Leu Ala Ser
            20                  25                  30

His Pro His Ser Leu Leu Thr Gly Thr Gln Phe Ser Gly Gln Thr Gln
        35                  40                  45

Ala Leu Ala Pro Ser Trp Cys Ala Xaa
    50                  55
```

Asn Glu Asn Gln Val Val Phe Ala Arg Val Asp Cys Asp Gln His Ser
85 90 95

Asp Ile Ala Gln Arg Tyr Arg Ile Ser Lys Tyr Pro Thr Leu Lys Leu
100 105 110

Phe Arg Asn Gly Met Met Met Lys Arg Glu Tyr Arg Gly Gln Arg Ser
115 120 125

Val Lys Ala Leu Ala Asp Tyr Ile Arg Gln Gln Lys Ser Asp Pro Ile
130 135 140

Gln Glu Ile Arg Asp Leu Ala Glu Ile Thr Thr Leu Asp Arg Ser Lys
145 150 155 160

Arg Asn Ile Ile Gly Tyr Phe Glu Gln Lys Asp Ser Asp Asn Tyr Arg
165 170 175

Val Phe Glu Arg Val Ala Asn Ile Leu His Asp Asp Cys Ala Phe Leu
180 185 190

Ser Ala Phe Gly Asp Val Ser Lys Pro Glu Arg Tyr Ser Gly Asp Asn
195 200 205

Ile Ile Tyr Lys Pro Pro Gly His Ser Ala Pro Asp Met Val Tyr Leu
210 215 220

Gly Ala Met Thr Asn Phe Asp Val Thr Tyr Asn Trp Ile Gln Asp Lys
225 230 235 240

Cys Val Pro Leu Val Arg Glu Ile Thr Phe Glu Asn Gly Glu Glu Leu
245 250 255

Thr Glu Glu Gly Leu Pro Phe Leu Ile Leu Phe His Met Lys Glu Asp
260 265 270

Thr Glu Ser Leu Glu Ile Phe Gln Asn Glu Val Ala Arg Gln Leu Ile
275 280 285

Ser Glu Lys Gly Thr Ile Asn Phe Leu His Ala Asp Cys Asp Lys Phe
290 295 300

Arg His Pro Leu Leu His Ile Gln Lys Thr Pro Ala Asp Cys Pro Val
305 310 315 320

Ile Ala Ile Asp Ser Phe Arg His Met Tyr Val Phe Gly Asp Phe Lys
325 330 335

Asp Val Leu Ile Pro Gly Lys Leu Lys Gln Phe Val Phe Asp Leu His
340 345 350

Ser Gly Lys Leu His Arg Glu Phe His His Gly Pro Asp Pro Thr Asp
355 360 365

Thr Ala Pro Gly Glu Gln Ala Gln Asp Val Ala Ser Ser Pro Pro Glu
370 375 380

Ser Ser Phe Gln Lys Leu Ala Pro Ser Glu Tyr Arg Tyr Thr Leu Leu
385 390 395 400

(2) INFORMATION FOR SEQ ID NO: 75:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

Met Ala Gly Ile His Arg Ala Phe Leu Val Phe Cys Leu Trp Gly Leu
1 5 10 15

Xaa Leu Cys Val Val Gly Gly Pro Trp Xaa
20 25

(2) INFORMATION FOR SEQ ID NO: 76:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

Met Ser Phe Ser Ser Pro Lys Ser Leu Leu Ser Leu Ile Ser Xaa
1 5 10 15

(2) INFORMATION FOR SEQ ID NO: 77:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 77:

Met Thr Ile Trp Gln Leu Phe Ala Val Leu Ile Val Leu Phe Ala Lys
1 5 10 15

Ser Arg Glu Ile Ser Thr Glu Gly Glu Pro Cys Val Leu Ser Lys Asn
20 25 30

Xaa

(2) INFORMATION FOR SEQ ID NO: 78:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

Met Leu Asn Pro Phe Xaa Gln Leu Leu Leu Val Leu Leu Phe Pro Glu
1 5 10 15

Trp Pro Thr Pro Leu His Xaa

20

(2) INFORMATION FOR SEQ ID NO: 79:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 173 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

```
Met Lys Thr Leu Phe Leu Gly Val Thr Leu Gly Leu Ala Ala Ala Leu
 1               5                  10                  15

Ser Xaa Thr Leu Xaa Glu Glu Asp Ile Thr Gly Thr Trp Tyr Val Lys
            20                  25                  30

Ala Met Val Val Asp Lys Thr Phe Arg Arg Gln Glu Ala Gln Lys Val
        35                  40                  45

Ser Pro Val Lys Val Thr Ala Leu Gly Gly Gly Lys Leu Glu Ala Thr
    50                  55                  60

Phe Thr Phe Met Arg Glu Asp Arg Cys Ile Gln Lys Lys Ile Leu Xaa
65                  70                  75                  80

Arg Lys Thr Glu Glu Pro Gly Lys Tyr Ser Ala Cys Glu Pro Leu Pro
                85                  90                  95

His Ser His Pro His Xaa Pro Pro Pro Pro Thr Pro Val His Gln Pro
            100                 105                 110

Pro Gln Val Glu Ser Ala Gln Ala Ala Leu Leu Pro Gly Pro Gln Leu
        115                 120                 125

Cys Pro Pro Pro Arg Arg Gly Trp Pro Leu Leu Pro Gly Gly Leu Val
    130                 135                 140

Ala Leu Thr Ser Asp Thr Gly Cys Asp Arg Leu Val Arg Ser Arg Asp
145                 150                 155                 160

Gly Pro Asp His Ala Cys Pro Leu Gly Gly Pro Ser His
                165                 170
```

(2) INFORMATION FOR SEQ ID NO: 80:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 208 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

```
Met Ala Asp Ser Ser Tyr Thr Ser Glu Val Gln Ala Ile Leu Ala Phe
 1               5                  10                  15

Leu Ser Leu Gln Arg Thr Gly Ser Gly Gly Pro Gly Asn His Pro His
            20                  25                  30
```

Gly Pro Asp Ala Ser Ala Glu Gly Leu Asn Pro Tyr Gly Leu Val Ala
35 40 45

Pro Arg Phe Gln Arg Lys Phe Lys Ala Lys Gln Leu Thr Pro Arg Ile
50 55 60

Leu Glu Ala His Gln Asn Val Ala Gln Leu Ser Leu Ala Glu Ala Gln
65 70 75 80

Leu Arg Phe Ile Gln Ala Trp Gln Ser Leu Pro Asp Phe Gly Ile Ser
85 90 95

Tyr Val Met Val Arg Phe Lys Gly Ser Arg Lys Asp Glu Ile Leu Gly
100 105 110

Ile Ala Asn Asn Arg Leu Ile Arg Ile Asp Leu Ala Val Gly Asp Val
115 120 125

Val Lys Thr Trp Arg Phe Ser Asn Met Arg Gln Trp Asn Val Asn Trp
130 135 140

Asp Ile Arg Xaa Val Ala Ile Glu Phe Asp Glu His Ile Asn Val Ala
145 150 155 160

Phe Ser Cys Val Ser Ala Ser Cys Arg Ile Val His Glu Tyr Ile Gly
165 170 175

Gly Tyr Ile Phe Leu Ser Thr Arg Glu Xaa Ala Arg Gly Glu Glu Leu
180 185 190

Asp Glu Asp Leu Phe Leu Gln Leu Thr Gly Gly His Glu Ala Phe Xaa
195 200 205

(2) INFORMATION FOR SEQ ID NO: 81:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 43 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

Met Ile Phe Leu Leu Phe Leu Thr Pro Leu Trp Leu Gln Lys Gly Ser
1 5 10 15

Ala Gly Lys Met Ser Gly Glu Phe Leu Tyr Ala Ser Leu Phe Gln Trp
20 25 30

Asn Tyr Phe Trp Arg Asn Lys Lys Val Cys Xaa
35 40

(2) INFORMATION FOR SEQ ID NO: 82:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 146 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

```
Met Pro Ser Gly Phe Gln Thr Cys Leu Leu Phe Thr Leu Ser Pro Phe
 1               5                  10                  15

Ser Leu Ser Lys Ile Val Gly Val Pro Ser Gln Gln Leu Pro Gly Gln
            20                  25                  30

Leu Ser Glu Gln Gly Gly Leu Cys Gly His Glu Gly Glu Pro Ala Arg
        35                  40                  45

Thr Val Pro Glu Thr Gln Leu Pro Leu Pro Phe Asn Ser Ala Gly Pro
    50                  55                  60

Pro His Leu Lys Cys Thr Gly Ala Gly Lys Arg Val Trp Ser Pro Pro
65                  70                  75                  80

Arg Arg Ala Ala Gln Glu Val Ser Leu Gln Leu Val Ser Cys Xaa Pro
                85                  90                  95

Cys Arg Gln Xaa Thr Ser Arg Ala Phe Ser Leu Ala Thr Asp Arg Thr
            100                 105                 110

Ala Ser Ala Arg Val Cys Cys Arg Phe Pro Phe Lys His Thr His Ser
        115                 120                 125

Pro His Pro Arg Arg Pro Glu Val Gln Gly Ala Trp Ala Val Val Pro
    130                 135                 140

Leu Xaa
145
```

(2) INFORMATION FOR SEQ ID NO: 83:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 83:

```
Met Pro Trp Arg Arg Ala Gly Leu Met Met Leu Pro Ile Ile Thr Gly
 1               5                  10                  15

Cys Cys Pro Cys Ser Ala Ser Ile Xaa
            20                  25
```

(2) INFORMATION FOR SEQ ID NO: 84:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

```
Met Lys Thr Leu Phe Leu Gly Val Thr Leu Gly Leu Ala Leu Pro Cys
 1               5                  10                  15

Pro Ser Pro Trp Xaa Arg Arg Ile Ser Gln Gly Pro Gly Thr Xaa
            20                  25                  30
```

(2) INFORMATION FOR SEQ ID NO: 85:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 374 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

```
Met Ser Val Pro Ala Phe Ile Asp Ile Ser Glu Glu Asp Gln Ala Ala
 1               5                  10                  15

Glu Leu Arg Ala Tyr Leu Lys Ser Lys Gly Ala Glu Ile Ser Glu Glu
            20                  25                  30

Asn Ser Glu Gly Gly Leu His Val Asp Leu Ala Gln Ile Ile Glu Ala
        35                  40                  45

Cys Asp Val Cys Leu Lys Glu Asp Asp Lys Asp Val Glu Ser Val Met
    50                  55                  60

Asn Ser Val Val Ser Leu Leu Leu Ile Leu Glu Pro Asp Lys Gln Glu
65                  70                  75                  80

Ala Leu Ile Glu Ser Leu Cys Glu Lys Leu Val Lys Phe Arg Glu Gly
                85                  90                  95

Glu Arg Pro Ser Leu Arg Leu Gln Leu Leu Ser Asn Leu Phe His Gly
            100                 105                 110

Met Asp Lys Asn Thr Pro Val Arg Tyr Thr Val Tyr Cys Ser Leu Ile
        115                 120                 125

Lys Val Ala Ala Ser Cys Gly Ala Ile Gln Tyr Ile Pro Thr Glu Leu
    130                 135                 140

Asp Gln Val Arg Lys Trp Ile Ser Asp Trp Asn Leu Thr Thr Glu Lys
145                 150                 155                 160

Lys His Thr Leu Leu Arg Leu Leu Tyr Glu Ala Leu Val Asp Cys Lys
                165                 170                 175

Lys Ser Asp Ala Ala Ser Lys Val Met Val Glu Leu Leu Gly Ser Tyr
            180                 185                 190

Thr Glu Asp Asn Ala Ser Gln Ala Arg Val Asp Ala His Arg Cys Ile
        195                 200                 205

Val Arg Ala Leu Lys Asp Pro Asn Ala Phe Leu Phe Asp His Leu Leu
    210                 215                 220
```

```
Thr Leu Lys Pro Val Lys Phe Leu Glu Gly Glu Leu Ile His Asp Leu
225                 230                 235                 240

Leu Thr Ile Phe Val Ser Ala Lys Leu Ala Ser Tyr Val Lys Phe Tyr
                245                 250                 255

Gln Asn Asn Lys Asp Phe Ile Asp Ser Leu Gly Leu Leu His Glu Gln
            260                 265                 270

Asn Met Ala Lys Met Arg Leu Leu Thr Phe Met Gly Met Ala Val Glu
        275                 280                 285

Asn Lys Glu Ile Ser Phe Asp Thr Met Gln Gln Glu Leu Gln Ile Gly
    290                 295                 300

Ala Asp Asp Val Glu Ala Phe Val Ile Asp Ala Val Arg Thr Lys Met
305                 310                 315                 320

Val Tyr Cys Lys Ile Asp Gln Thr Gln Arg Lys Val Val Val Ser His
                325                 330                 335

Ser Thr His Arg Thr Phe Gly Lys Gln Gln Trp Gln Gln Leu Tyr Asp
            340                 345                 350

Thr Leu Asn Ala Trp Lys Gln Asn Leu Asn Lys Val Lys Asn Ser Leu
        355                 360                 365

Leu Ser Leu Ser Asp Thr
    370
```

(2) INFORMATION FOR SEQ ID NO: 86:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

```
Met Ser Val Pro Ala Phe Ile Asp Ile Ser Glu Glu Asp
  1               5                  10
```

(2) INFORMATION FOR SEQ ID NO: 87:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 87:

```
Gln Ala Ala Glu Leu Arg Ala Tyr Leu Lys Ser Lys Gly Ala Glu
  1               5                  10                  15
```

(2) INFORMATION FOR SEQ ID NO: 88:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 88:

Ile Ser Glu Glu Asn Ser Glu Gly Gly Leu His Val Asp Leu Ala Gln
1 5 10 15

Ile

(2) INFORMATION FOR SEQ ID NO: 89:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 89:

Ile Glu Ala Cys Asp Val Cys Leu Lys Glu Asp Asp Lys Asp Val Glu
1 5 10 15

Ser Val

(2) INFORMATION FOR SEQ ID NO: 90:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 90:

Val Ala Arg Pro Ser Ser Leu Phe Arg Ser Ala Trp Ser Cys Glu Trp
1 5 10 15

(2) INFORMATION FOR SEQ ID NO: 91:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 91:

Leu Arg Leu Gln Leu Leu Ser Asn Leu Phe His Gly
1 5 10

(2) INFORMATION FOR SEQ ID NO: 92:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 92:

```
Lys Asp Val Glu Ser Val Met Asn Ser Val Val Ser Leu Leu Leu Ile
  1               5                  10                  15

Leu
```

(2) INFORMATION FOR SEQ ID NO: 93:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 93:

```
Asp Ala Ala Ser Lys Val Met Val Glu Leu Leu Gly Ser Tyr Thr Glu
  1               5                  10                  15

Asp Asn Ala Ser Gln Ala Arg Val Asp Ala
             20                  25
```

(2) INFORMATION FOR SEQ ID NO: 94:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 94:

```
Val Glu Ala Phe Val Ile Asp Ala Val Arg
  1               5                  10
```

(2) INFORMATION FOR SEQ ID NO: 95:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 95:

```
Met Ser Glu Ile Tyr Leu Arg Cys Gln Asp Glu Gln Gln Tyr Ala Arg
  1               5                  10                  15

Trp Met Ala Gly Cys Arg Leu Ala Ser Lys Gly Arg Thr Met Ala Asp
             20                  25                  30

Ser Ser Tyr
         35
```

(2) INFORMATION FOR SEQ ID NO: 96:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 96:

```
Leu Val Ala Pro Arg Phe Gln Arg Lys Phe Lys Ala Lys Gln Leu Thr
 1               5                  10                  15

Pro Arg Ile Leu Glu Ala His Gln Asn Val Ala Gln Leu Ser Leu Ala
            20                  25                  30

Glu Ala Gln Leu Arg Phe Ile Gln Ala Trp Gln Ser Leu
        35                  40                  45
```

(2) INFORMATION FOR SEQ ID NO: 97:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 97:

```
Val Gly Asp Val Val Lys Thr Trp Arg Phe Ser Asn Met Arg Gln Trp
 1               5                  10                  15

Asn Val Asn Trp Asp Ile Arg
            20
```

(2) INFORMATION FOR SEQ ID NO: 98:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 98:

```
Glu Glu Ile Asp Cys Thr Glu Glu Glu Met Met Val Phe Ala Ala Leu
 1               5                  10                  15

Gln Tyr His Ile Asn Lys Leu Ser Gln Ser
            20                  25
```

(2) INFORMATION FOR SEQ ID NO: 99:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 99:

Glu Glu Ile Asp Cys Thr Glu Glu Glu Met Met Val Phe Ala Ala Leu
1 5 10 15

Gln Tyr His Ile Asn Lys Leu Ser Gln Ser
20 25

(2) INFORMATION FOR SEQ ID NO: 100:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 100:

Lys Glu Leu Ser Phe Ala Arg Ile Lys Ala Val Glu Cys Val Glu Ser
1 5 10 15

Thr Gly Arg His Ile Tyr Phe Thr Leu Val
20 25

(2) INFORMATION FOR SEQ ID NO: 101:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 101:

Gly Trp Asn Ala Gln Ile Thr Leu Gly Leu Val Lys Phe Lys Asn Gln
1 5 10 15

Gln

(2) INFORMATION FOR SEQ ID NO: 102:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 102:

Leu Val Leu Gly Leu Ser Xaa Leu Asn Asn Ser Tyr Asn Phe Ser Phe
1 5 10 15

(2) INFORMATION FOR SEQ ID NO: 103:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 103:

```
His Val Val Ile Gly Ser Gln Ala Glu Glu Gly Gln Tyr Ser Leu Asn
  1               5                  10                  15

Phe
```

(2) INFORMATION FOR SEQ ID NO: 104:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 104:

```
His Asn Cys Asn Asn Ser Val Pro Gly Lys Glu His Pro Phe Asp Ile
  1               5                  10                  15

Thr Val Met
```

(2) INFORMATION FOR SEQ ID NO: 105:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 105:

```
Phe Ile Lys Tyr Val Leu Ser Asp Lys Glu Lys Lys Val Phe Gly Ile
  1               5                  10                  15

Val
```

(2) INFORMATION FOR SEQ ID NO: 106:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 106:

```
Ile Pro Met Gln Val Leu Ala Asn Val Ala Tyr Ile Ile
  1               5                  10
```

(2) INFORMATION FOR SEQ ID NO: 107:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 107:

```
Ile Pro Met Gln Val Leu Ala Asn Val Ala Tyr Ile Ile
 1               5                  10
```

(2) INFORMATION FOR SEQ ID NO: 108:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 108:

```
Asp Gly Lys Val Ala Val Asn Leu Ala Lys Leu Lys Leu Phe Arg
 1               5                  10                  15
```

(2) INFORMATION FOR SEQ ID NO: 109:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 109:

```
Ile Arg Glu Lys Asn Pro Asp Gly Phe Leu Ser Ala Ala
 1               5                  10
```

(2) INFORMATION FOR SEQ ID NO: 110:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 110:

```
Met Met Phe Gly Gly Tyr Glu Thr Ile
 1               5
```

(2) INFORMATION FOR SEQ ID NO: 111:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 111:

```
Tyr Arg Asp Glu Ser Ser Ser Glu Leu Ser Val Asp Ser Glu Val Glu
```

```
1               5                   10                  15

Phe Gln Leu Tyr Ser Gln Ile His
            20
```

(2) INFORMATION FOR SEQ ID NO: 112:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 136 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 112:

```
Tyr Ala Gln Asp Leu Asp Asp Val Ile Arg Glu Glu Glu His Glu Glu
  1               5                  10                  15

Lys Asn Ser Gly Asn Ser Glu Ser Ser Ser Ser Lys Pro Asn Gln Lys
             20                  25                  30

Lys Leu Ile Val Leu Ser Asp Ser Glu Val Ile Gln Leu Ser Asp Gly
         35                  40                  45

Ser Glu Val Ile Thr Leu Ser Asp Glu Asp Ser Ile Tyr Arg Cys Lys
     50                  55                  60

Gly Lys Asn Val Arg Val Gln Ala Gln Glu Asn Ala His Gly Leu Ser
 65                  70                  75                      80

Ser Ser Leu Gln Ser Asn Glu Leu Val Asp Lys Lys Cys Lys Ser Asp
                 85                  90                      95

Ile Glu Lys Pro Lys Ser Glu Glu Arg Ser Gly Val Ile Arg Glu Val
            100                 105                 110

Met Ile Ile Glu Val Ser Ser Ser Glu Glu Glu Glu Ser Thr Ile Ser
        115                 120                 125

Glu Gly Asp Asn Val Glu Ser Trp
    130                 135
```

(2) INFORMATION FOR SEQ ID NO: 113:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 113:

```
Met Leu Leu Gly Cys Glu Val Asp Asp Lys Asp Asp Asp Ile Leu Leu
  1               5                  10                  15

Asn Leu Val Gly Cys Glu Asn Ser Val Thr Glu Gly Glu Asp Gly Ile
             20                  25                  30

Asn Trp Ser Ile Ser
         35
```

(2) INFORMATION FOR SEQ ID NO: 114:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 114:

```
Asp Lys Asp Ile Glu Ala Gln Ile Ala Asn Asn Arg Thr Pro Gly Arg
 1               5                  10                  15

Trp Thr
```

(2) INFORMATION FOR SEQ ID NO: 115:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 115:

```
Gln Arg Tyr Tyr Ser Ala Asn Lys Asn Ile Ile Cys Arg Asn Cys Asp
 1               5                  10                  15

Lys Arg Gly His Leu Ser Lys Asn Cys Pro Leu Pro Arg Lys Val
            20                  25                  30
```

(2) INFORMATION FOR SEQ ID NO: 116:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 179 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 116:

```
Arg Arg Cys Phe Leu Cys Ser Arg Arg Gly His Leu Leu Tyr Ser Cys
 1               5                  10                  15

Pro Ala Pro Leu Cys Glu Tyr Cys Pro Val Pro Lys Met Leu Asp His
            20                  25                  30

Ser Cys Leu Phe Arg His Ser Trp Asp Lys Gln Cys Asp Arg Cys His
        35                  40                  45

Met Leu Gly His Tyr Thr Asp Ala Cys Thr Glu Ile Trp Arg Gln Tyr
    50                  55                  60

His Leu Thr Thr Lys Pro Gly Pro Pro Lys Lys Pro Lys Thr Pro Ser
65                  70                  75                  80

Arg Pro Ser Ala Leu Ala Tyr Cys Tyr His Cys Ala Gln Lys Gly His
                85                  90                  95
```

Tyr Gly His Glu Cys Pro Glu Arg Glu Val Tyr Asp Pro Ser Pro Val
100 105 110

Ser Pro Phe Ile Cys Tyr Tyr Xaa Asp Lys Tyr Glu Ile Gln Glu Arg
115 120 125

Glu Lys Arg Leu Lys Gln Lys Ile Lys Val Xaa Lys Lys Asn Gly Val
130 135 140

Ile Pro Glu Pro Ser Lys Leu Pro Tyr Ile Lys Ala Ala Asn Glu Asn
145 150 155 160

Pro His His Asp Ile Arg Lys Gly Arg Ala Ser Trp Lys Ser Asn Arg
165 170 175

Trp Pro Gln

(2) INFORMATION FOR SEQ ID NO: 117:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 117:

Leu Ser Ile Ile Phe Leu Ala Phe Val Ser Ile Asp Arg Cys Leu Gln
1 5 10 15

Leu

(2) INFORMATION FOR SEQ ID NO: 118:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 67 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 118:

Gly Ser Cys Phe Ala Thr Trp Ala Phe Ile Gln Lys Asn Thr Asn His
1 5 10 15

Arg Cys Val Ser Ile Tyr Leu Ile Asn Leu Leu Thr Ala Asp Phe Leu
20 25 30

Leu Thr Leu Ala Leu Pro Val Lys Ile Val Val Asp Leu Gly Val Ala
35 40 45

Pro Trp Lys Leu Lys Ile Phe His Cys Gln Val Thr Ala Cys Leu Ile
50 55 60

Tyr Ile Asn
65

(2) INFORMATION FOR SEQ ID NO: 119:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 119:

```
Ala Pro Leu Glu Thr Met Gln Asn Lys Pro Arg Ala Pro Gln Lys Arg
 1               5                  10                  15

Ala Leu Pro Phe Pro Glu Leu Glu Leu Arg Asp Tyr Ala Ser Val Leu
            20                  25                  30

Thr Arg Tyr Ser Leu Gly Leu Arg Asn Lys Glu Pro Ser Leu Gly His
        35                  40                  45

Arg Trp Gly Thr Gln Lys Leu Gly Arg Ser Pro Cys
    50                  55                  60
```

(2) INFORMATION FOR SEQ ID NO: 120:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 166 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 120:

```
Asn Arg Glu Arg Gly Gly Ala Gly Ala Thr Phe Glu Cys Asn Ile Cys
 1               5                  10                  15

Leu Glu Thr Ala Arg Glu Ala Val Val Ser Val Cys Gly His Leu Tyr
            20                  25                  30

Cys Trp Pro Cys Leu His Gln Trp Leu Glu Thr Arg Pro Glu Arg Gln
        35                  40                  45

Glu Cys Pro Val Cys Lys Ala Gly Ile Ser Arg Glu Lys Val Val Pro
    50                  55                  60

Leu Tyr Gly Arg Gly Ser Gln Lys Pro Gln Asp Pro Arg Leu Lys Thr
65                  70                  75                  80

Pro Pro Arg Pro Gln Gly Gln Arg Pro Ala Pro Glu Ser Arg Gly Gly
                85                  90                  95

Phe Gln Pro Phe Gly Asp Thr Gly Gly Phe His Phe Ser Phe Gly Val
            100                 105                 110

Gly Ala Phe Pro Phe Gly Phe Phe Thr Thr Val Phe Asn Ala His Glu
        115                 120                 125

Pro Phe Arg Arg Gly Thr Gly Val Asp Leu Gly Gln Gly His Pro Ala
    130                 135                 140

Ser Ser Trp Gln Asp Ser Leu Phe Leu Phe Leu Ala Ile Phe Phe Phe
145                 150                 155                 160

Phe Trp Leu Leu Ser Ile
                165
```

# INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis)*

**A.** The indications made below relate to the microorganism referred to in the description on page 29 , line N/A

**B. IDENTIFICATION OF DEPOSIT** Further deposits are identified on an additional sheet ☐

Name of depositary institution
American Type Culture Collection

Address of depositary institution (*including postal code and country*)

10801 University Boulevard
Manassas, Virginia 20110-2209
United States of America

| Date of deposit May 22, 1997 | Accession Number 209075 |
|---|---|

**C. ADDITIONAL INDICATIONS** *(leave blank if not applicable)* This information is continued on an additional sheet ☐

**D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE** *(if the indications are not for all designated States)*

EUROPE

In respect of those designations in which a European Patent is sought a sample of the deposited microorganism will be available until the publication of the mention of the grant of the European patent or until the date on which the application has been refused or withdrawn or is deemed to be withdrawn, only by the issue of such a sample to an expert nominated by the person requesting the sample (Rule 28(4)EPC).

**E. SEPARATE FURNISHING OF INDICATIONS** *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later (*specify the general nature of the indications, e.g., "Accession Number of Deposit"*)

| For receiving Office use only | For International Bureau use only |
|---|---|
| ☑ This sheet was received with the international application | ☐ This sheet was received by the International Bureau on: |
| Authorized officer JERYL McDOWELL 703-305-3639 | Authorized officer |

# INDICATIONS RELATING TO A DEPOSITED MIC~~ROORGANISM~~

(PCT Rule 13*bis)*

**A.** The indications made below relate to the microorganism referred to in the description on page 30, line N/A

**B. IDENTIFICATION OF DEPOSIT** Further deposits are identified on an additional sheet ☐

Name of depositary institution
American Type Culture Collection

Address of depositary institution (*including postal code and country*)

10801 University Boulevard
Manassas, Virginia 20110-2209
United States of America

| Date of deposit May 8, 1997 | Accession Number 209022 |
|---|---|

**C. ADDITIONAL INDICATIONS** *(leave blank if not applicable)* This information is continued on an additional sheet ☐

**D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE** *(if the indications are not for all designated States)*

EUROPE

In respect of those designations in which a European Patent is sought a sample of the deposited microorganism will be available until the publication of the mention of the grant of the European patent or until the date on which the application has been refused or withdrawn or is deemed to be withdrawn, only by the issue of such a sample to an expert nominated by the person requesting the sample (Rule 28(4)EPC).

**E. SEPARATE FURNISHING OF INDICATIONS** *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later (*specify the general nature of the indications, e.g., "Accession Number of Deposit"*)

| For receiving Office use only | For International Bureau use only |
|---|---|
| ☑ This sheet was received with the international application | ☐ This sheet was received by the International Bureau on: |
| Authorized officer<br>JERYL McDOWELL<br>703-305-3639 | Authorized officer |

*What Is Claimed Is:*

1. An isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 95% identical to a sequence selected from the group consisting of:

(a) a polynucleotide fragment of SEQ ID NO:X or a polynucleotide fragment of the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;

(b) a polynucleotide encoding a polypeptide fragment of SEQ ID NO:Y or a polypeptide fragment encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;

(c) a polynucleotide encoding a polypeptide domain of SEQ ID NO:Y or a polypeptide domain encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;

(d) a polynucleotide encoding a polypeptide epitope of SEQ ID NO:Y or a polypeptide epitope encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;

(e) a polynucleotide encoding a polypeptide of SEQ ID NO:Y or the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X, having biological activity;

(f) a polynucleotide which is a variant of SEQ ID NO:X;

(g) a polynucleotide which is an allelic variant of SEQ ID NO:X;

(h) a polynucleotide which encodes a species homologue of the SEQ ID NO:Y;

(i) a polynucleotide capable of hybridizing under stringent conditions to any one of the polynucleotides specified in (a)-(h), wherein said polynucleotide does not hybridize under stringent conditions to a nucleic acid molecule having a nucleotide sequence of only A residues or of only T residues.

2. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding a secreted protein.

3. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding the sequence identified as SEQ ID NO:Y or the polypeptide encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X.

4. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises the entire nucleotide sequence of SEQ ID NO:X or the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X.

5. The isolated nucleic acid molecule of claim 2, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.

6. The isolated nucleic acid molecule of claim 3, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.

7. A recombinant vector comprising the isolated nucleic acid molecule of claim 1.

8. A method of making a recombinant host cell comprising the isolated nucleic acid molecule of claim 1.

9. A recombinant host cell produced by the method of claim 8.

10. The recombinant host cell of claim 9 comprising vector sequences.

11. An isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence selected from the group consisting of:

(a) a polypeptide fragment of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

(b) a polypeptide fragment of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z, having biological activity;

(c) a polypeptide domain of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

(d) a polypeptide epitope of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

(e) a secreted form of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

(f) a full length protein of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

(g) a variant of SEQ ID NO:Y;

(h) an allelic variant of SEQ ID NO:Y; or

(i) a species homologue of the SEQ ID NO:Y.

12. The isolated polypeptide of claim 11, wherein the secreted form or the full length protein comprises sequential amino acid deletions from either the C-terminus or the N-terminus.

13. An isolated antibody that binds specifically to the isolated polypeptide of claim 11.

14. A recombinant host cell that expresses the isolated polypeptide of claim 11.

15. A method of making an isolated polypeptide comprising:

(a) culturing the recombinant host cell of claim 14 under conditions such that said polypeptide is expressed; and

(b) recovering said polypeptide.

16. The polypeptide produced by claim 15.

17. A method for preventing, treating, or ameliorating a medical condition, comprising administering to a mammalian subject a therapeutically effective amount of the polypeptide of claim 11 or the polynucleotide of claim 1.

18. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:

(a) determining the presence or absence of a mutation in the polynucleotide of claim 1; and

(b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or absence of said mutation.

19. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:

(a) determining the presence or amount of expression of the polypeptide of claim 11 in a biological sample; and

(b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or amount of expression of the polypeptide.

20. A method for identifying a binding partner to the polypeptide of claim 11 comprising:

(a) contacting the polypeptide of claim 11 with a binding partner; and

(b) determining whether the binding partner effects an activity of the polypeptide.

21. The gene corresponding to the cDNA sequence of SEQ ID NO:Y.

22. A method of identifying an activity in a biological assay, wherein the method comprises:

(a) expressing SEQ ID NO:X in a cell;

(b) isolating the supernatant;

(c) detecting an activity in a biological assay; and

(d) identifying the protein in the supernatant having the activity.

23. The product produced by the method of claim 22.